Building a Genetic System in Yeast

to Search for High Affinity Proteins in Sequence Space

Matthew Douglas Merguerian

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2013

ABSTRACT

Building a Genetic System in Yeast

to Search for High Affinity Proteins in Sequence Space

Matthew Douglas Merguerian

Binding proteins (both natural and man-made) have the ability to bind tightly and specifically with small molecules and other biopolymers. Binding proteins can function as human therapeutics, diagnostics, and as tools for scientific research. Given the wide range of potential applications, there is great interest in both academia and industry to develop methods for discovering novel binders. An important step in discovering new binders is called affinity maturation, when an initial hit that shows some ability to bind the target is further mutated in additional steps to improve binding affinity, specificity, solubility, pharmacokinetic profile. Ideally, the methods for affinity maturation would allow for cookbook protocols, be successful for arbitrary targets of interest, and be minimally resource intensive. Although traditional methods for affinity maturation have had some stunning successes over the past, the state of the field is still far from this ideal. In **Chapter 1**, I discuss the current state of the protein engineering field. In **Chapter 2**, I discuss the use of phenotypic selection for yeast-display protein binders, and test these systems on simple loop libraries.. In **Chapter 3**, I construct a genetic system in yeast that can mutate a protein loop via homologous recombination, and test its recombination function. In **Chapter 4**, I mate libraries that target two different loops, and run FACS on the combinatorial libraries. In **Chapter 5**, I discuss future directions for the project.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

5-FOA	5-fluoroorotic acid
A	deoxyadenosine
Amp^R	ampicillin resistance
A_x	absorbance at x nm
bp	base pair
°C	degrees Celsius
C	deoxycytoside
cAMP	adenosine 3′,5′-cyclic monophosphate
CDR	complementarity-determining region
CML	chronic myelogenous leukemia
ddH_2O	deionized, distilled H_2O
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
DSB	double-stranded break
E. coli	*Escherichia coli*
et al.	et alia
FACS	fluorescence-activated cell sorting
FITC	fluorescein isothiocyanate
10Fn3	10th human fibronectin type III domain
g	gram
G	deoxyguanosine
G418	geneticin
gal	galactose
GFP	green fluorescent protein
h	hour
HR	homologous recombination
HRS	heritable recombination system

IDT	Integrated DNA Technologies
IPTG	isopropyl □-D-thiogalactopyranoside
Kan^R	kanamycin resistance
kb	kilobase
K_d	equilibrium dissociation constant
L	liter
M	moles per liter
Mb	megabase
MCS	multiple cloning site
mg	milligram
min	minute
mL	milliliter
nm	nanometer
nM	nanomoles per liter
OD_x	optical density at x nm
oligo	oligonucleotide
ORF	open reading frame
ori	origin of replication
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PE	phycoerythrin
pM	picomoles per liter
RNA	ribonucleic acid
rpm	revolutions per minute
s (or sec)	second
S. cerevisiae	*Saccharomyces cerevisiae*
SC	synthetic complete
SDS	sodium dodecyl sulfate

t	time
T	deoxythymidine
T_M	melting temperature
UV	ultraviolet
VEGF	Vascular endothelial growth factor
xg	Relative Centrifugal Force (in multiples of g, earth's gravitational acceleration)
Y2H	yeast two-hybrid
YPD	yeast peptone dextrose media
µg	microgram
µL	microliter
µM	micromoles per liter

Acknowledgements

I thank the NIH NIGMS, both for the Medical Scientist Training Program that has provided me with funding that allowed me to carry out this research and also for the R01 funding that has supported our lab.

First and foremost, I thank my advisor, Professor Virginia Cornish. None of this would have been possible without her. She has *always* been excited about the science at every step of the way. Her frequent input and guidance has been invaluable to me. One of her most endearing qualities was her concern for her students as whole people, as she showed concern for our lives outside of the lab. She sets an extraordinary example of balancing work and family that I hope to one day emulate.

I thank the members of the Cornish Lab next. I am immeasurably grateful for their daily (really hourly) input. They were always supportive, during good times and bad. I cannot thank them enough for taking time out of their own busy schedules to teach me with thoroughness and patience. I thank Laura Wingler, who trained me on the basics, and who was a graduate student of the highest caliber. I thank Nili Ostrov, who continues to teach me to this day, who has been an excellent consult and a great friend in the lab. I thank Dante Romanini, who always had insightful advice on experiments and presentations; his solutions were often elegant and always practical. I thank Tong Shu Li, an undergraduate who worked on this project for two summers, who was very sharp, hard-working, and delightful to work with. And I thank all the other members, past and present, for their camaraderie and helpfulness.

I thank the members of my graduate committee, Professor Ronald Breslow and Professor Scott Banta, who generously took time out of their schedules to provide extremely stimulating discussion and helpful guidance at my 2nd year research defense and 4th year original research proposal. I also thank the additional members of my thesis committee, Professor Richard Friesner, Professor Eric Greene, and Professor Richard Baer for their time and helpful discussion.

I would like to thank Dr. Siu-hong Ho at the CCTI flow cytometry core at Columbia University Medical Center for helping me to run FACS sorting and flow cytometry experiments.

I would like to thank Shohei Koide, his wife Akiko, and John Wojcik who have been helpful collaborators in the present work. Shohei, who was my undergraduate advisor, took me in as a college freshman with absolutely no lab experience. My experience in the Koide Lab shaped how I think about science.

I thank my fiancée Carolyn Arena for being my light and motivation during graduate school, for her understanding and patience during my long nights at the lab, for taking care of me during the writing of this thesis and for being a second pair of eyes for editing.

I thank my parents who encouraged my interest in science as a child and who taught me how to work hard. Their unwavering love and support throughout my life has provided me with the strength to achieve my goals.

For Carolyn

Chapter 1

Searching for functional protein binders in sequence space

1.1 Abstract

Protein binders with high affinity and high specificity for their ligand are invaluable handles on the nanoscale world that allow us to investigate and manipulate biological systems with surgical precision. Current technologies for searching for these high-functioning binders in mutational sequence space still sample only an insignificant fraction of all possible variants. In this chapter, I present the idea for a genetic system in *Saccharomyces cerevisiae* capable of combinatorially walking through sequence space through repeated rounds of mutagenesis and selection *in vivo*. This genetic system brings together a number of different components which are introduced and discussed here. These include a heritable recombination system for cassette based mutagenesis (recently developed by our lab), the robust yeast display system for selection of high-affinity binders, the fibronectin-based monobody domain which serves as our scaffold for building binders, and the SH2 domains which serve as our target of binding.

1.2 Binding proteins

Proteins perform an overwhelmingly diverse set of functions within living organisms. They can catalyze highly specific chemical reactions. Some transport proteins are molecular motors that can ferry cellular cargo to and fro along the skeletal roadways of the cell. Other transport proteins are located in the cell membrane and function in homeostasis, bringing vital nutrients from the outside world into the cell's interior and regulating flux of ions. Some proteins are largely structural, providing shape to the cell and allowing cells to adhere to their environment[1]. And many proteins function in molecular recognition by binding to other molecules in the cell. The promise of protein engineering is that one day this vast array of functionality will be at our technological disposal.

The binding function of proteins is well understood at a basic level. Targets of binding, called ligands, can be essentially anything: informational polymers like DNA and RNA, other proteins (including

itself), small organic molecules like metabolites, and inorganic ions like Ca^{2+} or Fe^{2+}. Binding proteins can grab onto their ligands and hold them with a certain tightness, a quality which is called affinity. An experimentally determined parameter called the equilibrium constant of dissociation, or K_d, is very often used as the quantitative measure of affinity. If we think of two proteins A and B that can bind to each other in a 1:1 ratio, we can think of their interaction symbolically as an equilibrium between two reactions, the dissociation reaction of the complex and the association reaction of the individuals

$$A + B \underset{k_{off}}{\overset{k_{on}}{\rightleftarrows}} A \cdot B$$

The equilibrium dissociation constant then is the ratio of the kinetic rates for the two reactions.

$$K_d = \frac{k_{off}}{k_{on}}$$

In protein affinity maturation, the goal is generally to increase affinity, and thus decrease K_d. A protein's affinity for its ligand is a function of the geometric complementarity of the atoms at the binding interface, the chemical complementarity (e.g., acidic vs basic residues), and the flexibility of the protein at the interface[2,3]. A rough guideline is that K_d's in the 10^{-8}-10^{-9} M range (nanomolar or nM) are considered high affinity, K_d's in the 10^{-10} – 10^{-12} M (picomolar, or pM are considered very high affinity, and 10^{-12} M and below (femtomolar or fM) are considered extremely high affinity. Some proteins will be promiscuous, binding to many different proteins[4,5]. Others will be highly specific, binding strongly only to one or two other cellular components. The ability to discern one potential ligand from another makes proteins a handle on the molecular world, which makes them incredibly useful.

Antibodies are the most familiar class of binding proteins, and the wide variety of applications that have been found for them underscores the ever-increasing importance of affinity reagents[6]. Antibodies have come to be indispensable tools for scientific research. In immunoblotting, antibodies are used to specifically label a ligand of interest, which allows scientists to confirm the presence or absence of that ligand in the cells that they are studying[7]. Moreover, antibodies that are affixed to a solid substrate can be used to separate and purify the ligand from a complex molecular milieu, such as a cell lysate. In making a clinical diagnosis of disease, antibodies are used to stain specific molecular markers, which can

be visualized under a microscope by a pathologist[8]. This ability to read molecular signatures is becoming more important as clinicians search for molecular markers to tell apart diseases that appear to be grossly similar but that have different prognoses and therapeutic susceptibilities. For example, the estrogen, progesterone, and HER2 biomarkers are used to diagnose subtypes of breast cancer, which are then treated with tageted therapies (trastuzumab for HER2 positive, tamoxifen for estrogen receptor positive). Antibodies can also be used to activate or inhibit physiologic targets, such as cell-signaling receptors, and thus have found much application recently as therapeutic agents for cancer and autoimmune disorders[9–14]. These are all applications for which there already is and will increasingly be a demand for affinity reagents

Antibodies have a number of limitations as affinity reagents[15,16]. They are large in size and difficult to express at high-concentrations. The complicated structure also makes it extremely difficult to express antibodies recombinantly in different organisms that researchers might wish to study. Expression from eukayotic cell culture is necessary to have proper folding and glycosylation. Structurally, they are held together by multiple disulfide linkages, and so they must be kept under oxidizing conditions in order to function properly. Thus, despite being excellent handles on the molecular world, antibodies are not viable for studying the intracellular environment of living cells, where they might otherwise find a great deal of scientific application.

Given these considerations, many research groups have identified alternative scaffolds that are useful for building affinity reagents[15,17–20]. Alternative scaffolds are protein domains that have certain properties in common, such as a highly stable core that drives folding and solvent exposed regions that can tolerate a large amount of mutation. This allows a certain amount of reliability in using these domains as starting points for mutation and selection. Anytime you start a new binder evolution project, there is a well-defined region of the molecule to mutate (analogous to the hypervariable loops in antibodies[21]), and it can reasonably be expected that a large fraction of the resulting mutants will fold. Thus, there is some degree of structural separation of the two functions of the protein: residues responsible for driving folding are separate from those that are most likely to form a binding interface, which allows us to think of these domains as “scaffolds” upon which functionality can be built without worrying about changing the overall structure.

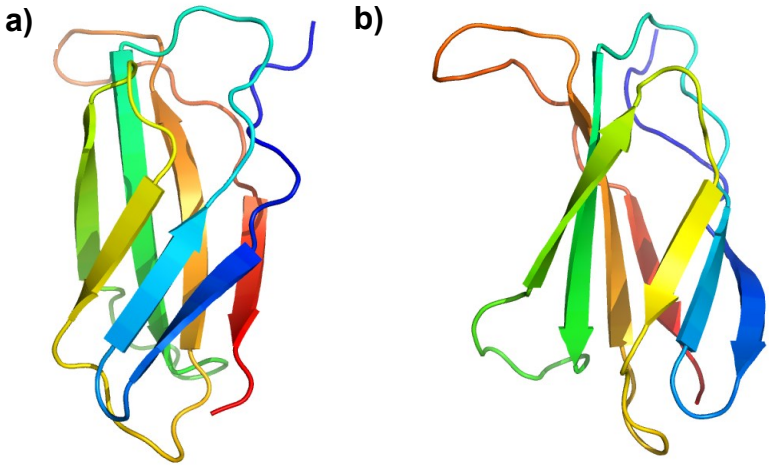


Figure 1-1. ^{10}Fn3 domain shown from two angles. β strands named alphabetically from A-G. The three loops commonly targeted for mutagenesis are the BC loop (light blue), the DE loop (yellow), and the FG loop (orange).
b) rotated ~60° relative to **a)**

The monobody scaffold (Figure 1-1), or ^{10}Fn3, is one of the most widely used of these alternative scaffolds, and it is the one we decided to use in the present study. ^{10}Fn3 was first utilized for directed evolution by Koide[22]. It is derived from the 94-residue 10th human fibronectin type III domain[23–25] and has an Ig-type fold structure[26]. It has favorable solubility[27] and stability[28,29] characteristics that allow it to be remarkably tolerant of mutation[30,31]. Monobodies have been used with phage display, yeast display, and mRNA display[24]. It has even found therapeutic application as CT-322, an anti-angiogenic agent[32–35]. It has successfully been used to target diverse targets, such as estrogen receptor, ySUMO, Abl1 SH2[36], MBP, epidermal growth factor receptor[37], VEGF receptor 2[38], and lysozyme[39].

1.3 Yeast display

In order to use directed evolution (see below) to search for high affinity monobodies, we require a method to select for the desired binding function. A number of robust display technologies are available that serve the purpose of connecting phenotype with genotype by physically linking each DNA molecule

with its corresponding protein variant. The most commonly used are yeast display and phage display[40–44]. mRNA and ribosome display[45–47] are also worthy of note for their very large library sizes but are technically challenging to work with.

The phenotypic selection that we use in our present work is based on the yeast display platform developed and described by Boder and Wittrup[48]. The basic idea (Figure 1-2) is that the protein scaffold to be diversified is translationally fused to a membrane protein, Aga2p that is expressed on the surface of yeast cells. The expression is inducible, since the fusion protein is under the control of the GAL1 promoter. When the cells are grown in the presence of galactose (induction), the fusion is expressed on the surface; when cells are grown in the presence of glucose, the expression is repressed. Boder et al. determined that there are about $3x10^4$ copies of the fusion protein displayed on the surface of the average galactose-induced yeast cell by comparing the fluorescence intensity of yeast displaying cells to reference standard microspheres.

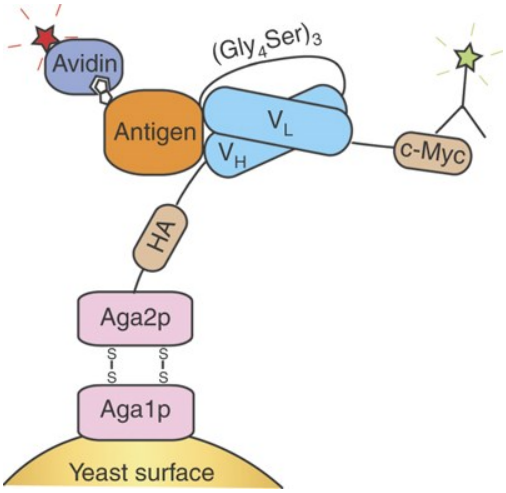


Figure 1-2. Schematic of yeast cell surface display, adapted from Chao et al. 2006. The scaffold of interest (in this case an scFv) is shown in blue, binding to its ligand in orange.

Yeast display is a robust and mature technology[44], which gave us a great deal of confidence when planning on incorporating it into our genetic system. Many aspects of yeast display have been optimized, and detailed protocols have been written for effectively implementing the platform for new targets[49–51]. Yeast display was used successfully to obtain a femtomolar binder[52], demonstrating its ability to make In a direct comparison between yeast display and phage display in targeting HIV-1 gp120,

more high-affinity clones were obtained from the yeast library[53]. It has been applied towards the affinity maturation of computational hits[54].

Yeast display allows two different, but complementary, selection modalities to be used in binder evolution, and both were exploited in for this monobody project. Ligand-coated magnetic beads enable high-throughput selection of binders[55]. In this selection, binding proteins on the surface of the yeast bind to ligand on the surface of the bead, linking the cell and the bead. A magnetic field is then applied, which pulls the beads to the side of the tube, along with any cells that are bound. Cells with non-functional binding variants remain in suspensions, and are then pipette off, leaving just the functional binders on the beads. As a result of avidity, even binders with modest affinity are retained. This is important in a combinatorial mutagenesis system such as ours, since a moderate functioning binder might actually go on to become a high-affinity binder after the following round of mutagenesis, and we would not like to remove its genes from the pool prematurely. Flourescence activated cell sorting is the second selection modality, and can be used for fine affinity discriminations between binders. Here, the ligand is covalently attached to a fluorophore. A yeast cell which expresses a higher affinity binder will have more fluorescent ligand molecules bound to its surface, and thus will have a more intense fluorescence signal. The intensity of fluorescent ligand signal can be controlled for expression level of the binder on the surface of the yeast cell, which crucially provides a way to distinguish between cells with high-affinity binders from cells with medium binders that are merely expressed at a very high copy number on the cell surface.

One of the additional benefits of yeast is that large libraries are readily stored and propagated. This means that a lot of effort can be front-loaded into the construction of a large library, which can then be used repeatedly for many different applications (in our case, cassette libraries targeting different) without having to build a new library. Feldhaus and colleagues demonstrated this with the construction of a 10^9 yeast scFv library built from PCR amplification of human variable region repertoire[56]. They showed that 5 rounds of propagation, each one involving passage of the cells, 100x cell amplification, and a -80 °C freeze-thaw cycle, did not substantially alter the prevalence of a particular clone of interest, indicating that there was low intrinsic growth bias when working with yeast display libraries.

1.4 Directed evolution of proteins

In order to find higher affinity monobodies, this project utilizes directed evolution, a technique in which the Cornish Lab has substantial prior experience[57–62]. Simply stated, directed evolution is any process involving repeated cycles of genetic mutation and phenotypic selection. In terms relevant for this current project, a gene sequence encoding a monobody (either the "wild-type" or some low-affinity binder) is used as a template to create a large number of different mutant sequences. Then, induction of yeast display results in the expression of monobody protein on the surface of the yeast. In the selection, cells expressing proteins with higher affinity
are physically separated from the rest. The cycle can then be repeated by taking the winning genetic material from this pool, mutating it again, and selecting for function again.

In directed evolution, the term "library" refers to the entire collection of variants that has been created via mutation. In yeast display, each variant copy of the gene is contained within a yeast cell, so a flask or tube containing many such yeast, each with a slightly different copy of the gene, is said to contain the library. There is no strong consensus in the literature with regard to terminology related to libraries. This leads to ambiguity, and it is worth making some quick definitions. I call "complexity" the number of unique sequences contained within a library. Following Bosley, I use "diversity" as a qualitative description of the average sequence difference between any two variants in the library[63]. Lastly, I use "quality" in the following sense: a library that has few premature truncations due to nonsense stop codons and has low redundancy would be said to be of higher quality than one in which 30% of the variants were truncations, and 20% were all copies of a single variant[64]. Some concerns about working with library complexity are addressed in Appendix I.

The number of possible protein mutations for a fixed polypeptide length is vast (and even larger if length is allowed to vary). For a rather small protein with length of 100 residues, the number of different residue sequence possibilities is $20^{100} = 1.27 \times 10^{130}$. This potential variability allows the set of all naturally observed proteins to have such incredibly diverse functions. In a theoretical sense, we'd like to find the globally optimal sequence for the phenotypic function we wish to evolve. The limitations of the physical world, however, prevent us from doing anything of this sort. The largest library sizes achievable with current technology are only 10^{12} - 10^{13} using mRNA display and ribosome display[46]. This means that in

the very best situation (even being generous with codon degeneracy and the practical need for oversampling the library), we can create libraries where at most 10 residues are fully randomized. In the context of binding proteins, this means that one of the three hypervariable loops of an antibody variable domain can be practically randomized and searched at a time. There is clearly a need to carefully choose which residues will be varied, and how they will be varied, in order to limit the search of sequence space to a subset of sequences that you can experimentally work with[65–72]. Library diversity is discussed further in Chapter 2.

1.5 Affinity maturation

Affinity maturation refers to a special type of directed evolution, where an initial hit with some affinity to the ligand of interest has been found, but further mutation is needed in order to obtain optimize obtain higher affinity or better specificity. Here, structural studies like X-ray crystallography and thermodynamic studies like alanine scanning mutagenesis can be used to provide some information about the binding interaction. This information can then guide further mutagenesis efforts, and thus there is more of a role for smart library design.

Why is higher affinity desired? Higher affinity binders are more potent, in the sense that a lower concentration of the binder is needed for the same effect. By example, if a clinician wanted to block a signaling receptor in the body with a therapeutic affinity reagent, a small amount of high affinity reagent could block the same number of receptors as a larger amount of medium affinity reagent. In this situation, poor affinity is detrimental for two reasons. First, higher concentrations of the therapeutic can lead to increased side-effects, both from allergic reactions of the body against the therapeutic and from increased interaction of the therapeutic with secondary targets. Secondly, protein therapeutics are currently very expensive to produce[16], so higher amounts of protein per dose adds substantial cost. The considerations are similar for affinity reagents in scientific research and diagnostics. An scFv scaffold binder being developed as a diagnostic for prostate-specific antigen was 10-fold more sensitive for its target after an 8.5 fold affinity increase[73]. A bifunctional antibody being developed as a radio-imaging label for tumor initially had an affinity of 10 nM for radionuclide chelates[74]. After an affinity maturation of

1000-fold to 8.2 pM, the antibody demonstrated a significantly increased retention of radionuclide chelate *in vivo*.

Our strategy for affinity maturing binders is based on combinatorial matching of loop libraries. Genentech provided an excellent demonstration of the power of this mutational strategy in the development of an anti-VEGF antibody that would ultimately become Lucentis, a therapeutic antibody for macular degeneration®[75]. The starting point was a monoclonal antibody with low nM affinity for VEGF. The Genentech team chose a targeted loop library approach, three libraries corresponding to heavy chain CDR loops and the fourth targeting a framework region. A crystal structure of the original binder and alanine scanning data were used to limit randomization to 3-5 residues so that the maximum theoretical complexity could be exhaustively searched. 7 rounds of enrichment were carried out with each library, sequencing after the final rounds to identify the winning mutations. Importantly, individual winner mutations from two of the loop libraries were then combined by site-directed mutagenesis to arrive at a binder with 22-fold affinity improvement over the starting point. With our technology we hope to recapitulate such a mutational strategy completely *in vivo*, simply by mating and selecting yeast.

1.6 SH2 domains

Knowing that we wanted to use combinatorial mutagenesis to affinity mature a monobody, we needed a starting monobody with high affinity to an interesting target. In 2010, our collaborators in the Koide Lab at the University of Chicago published the discovery of a monobody binder, called "HA4," that could bind to the SH2 domain of human Abl kinase (both Abl1 and Abl2) with a K_d = 7 nM[36]. This monobody was found to compete with phosphopeptide (SH2's physiological ligand), indicating that the monobody was binding to an epitope on SH2 important for phosphopeptide recognition. A crystal structure of HA4 in complex with Abl1 SH2 was obtained with 1.75 Å resolution. *In vivo* studies demonstrated that transfection of mammalian cells with HA4 could inhibit processive phosphorylation of paxillin protein (a demonstrated substrate of Abl[76]) and inhibit phosphorylation of STAT5, a endogenous downstream signaling target of Bcr-Abl[77,78].

Following up on this study, Wojcik and the Koide group developed a second monobody called 7c12 which inhibited Bcr-Abl in an *in vitro* kinase assay, but bound to the SH2 domain on the face opposite from where HA4 bound[79,80]. Capitalizing on this, they then connected the two monobodies covalently with a Gly-Ser linker hoping to increase affinity and specificity. Indeed, they found increased potency, with the HA4-7c12 monobody inhibiting Bcr-Abl *in vitro* at low micromolar concentrations. In CML K562 cells, expression of HA4-7c12 resulted in decreased phosphorylation of the Tyr412 in the activation loop of Bcr-Abl. The tandem monobody was found to have a pro-apoptotic effect on both CML cell lines and CML cells taken from patient, with a ~65% increase in apoptotic cells compared to non-binding control in chronic phase cells. This study validates the use of monobodies in probing cellular function and as tools for target validation for therapeutic development.

The HA4 monobody was attractive to us for a number of reasons. The SH2 domains are regulators of tyrosine kinases, which act as hubs within intricate signaling networks involved in myriad cellular functions[81,82]. Additionally, a crystal structure and alanine scanning results were obtained for HA4, which allowed us to think about loop library diversity within a structural context and provided useful information for troubleshooting the development of our genetic system. Importantly, the intermediate library of clones from which HA4 was functionally selected was available, and we took advantage of this to create library diversity as discussed in Chapter 4.

Also of interest, the whole SH2 domain family has been recognized as an important set of targets to challenge our ability to develop binders with specificity[83–86]. Many of the family members have close sequence and structural similarity. A recent paper reported the development of a panel of antibodies specific to 20 human SH2 domains using phage display[87]. The project was carried out in a high throughput manner, with an eye towards developing research tools for proteomics. Out of an initial library of 10^{10} different binders, 1292 phage hits were shown to bind antigen, with at least 150 unique binders demonstrating “monospecificity” for one of the SH2 domains, where monospecificity was defined as having an ELISA binding signal for the specific target > 10 the binding signal for any of the other 19 SH2 domains included in the panel. The 20 SH2 domains used in this panel (a subset of the 120 SH2 domains extant in the human genome) ranged in pairwise sequence similarity from 20-89% identity.

Specificity is a separate problem from affinity[2,4], but it is an extremely important one. The process of intentionally altering specificity through directed evolution is still being actively explored. Protein microarray data[36] showed that HA4 can bind to other SH2 domains, most noticeably at concentrations of HA4 at 32 nM. Although not the focus of the current work, it would be interesting to try and use combinatorial loop "walking"[88] to try to shift HA4's specificity for Abl SH2 to another domain (Chapter 5).

1.7 Harnessing the power of combinatorial mixing

In eukaryotes, sexual pairing of genes followed by homologous recombination during chromosomal crossover can result in genetic shuffling of a gene. This allows mutations that arose independently at different locations within two copies of the same gene to be combined into a single copy of that gene. This is a very successful strategy in natural evolution, as it obviates the need for two or more presumably rare mutations to arise sequentially within one ancestral lineage.

Inspired by this strategy, the Cornish lab has been interested in combining mating, homologous recombination, and phenotypic selection all into one system, which we call a "heritable recombination system" (HRS). Romanini and coworkers from our lab have recently reported the first application of HRS for combinatorially crossing the beneficial mutations from separate libraries[89]. This system had three major components each on a separate plasmid within a single yeast strain. First, it had a plasmid containing the gene to be mutated, the TRP1 biosynthetic gene. The second plasmid in the system is a cassette plasmid, which contains a mutational codon flanked by homology on both sides, which are further flanked by endonuclease cut sites (Figure 1-3b). The third and final plasmid contained the I-SceI endonuclease enzyme.

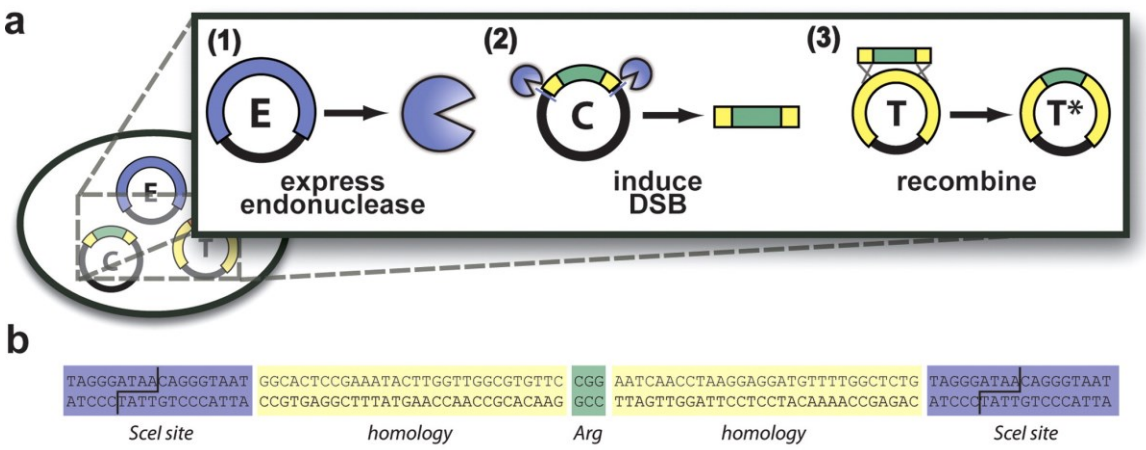


Figure 1-3. Homologous recombination of mutagenic cassette into target gene stimulated by endonuclease (reproduced from Fig. 2, Romanini 2012[89]). **a,** A general schematic for induction of mutagenesis. The endonuclease, shown as a blue pac-man, is expressed after growth in galactose and introduces DSB's into the cassette plasmid "C". The cut ends of the cassette stimulate recombination with the homologous portion of the TRP1 gene (yellow) in the target plasmid "T". **b,** A close-up of the mutagenic cassette contained within the cassette plasmid.

The basic logic of the design (Figure 1-3a) is that the endonuclease is under the control of the GAL1 promoter, which has a tight repression of expression when the yeast cells are in glucose, and strong activation of expression in galactose. By inducing the cells in galactose, I-SceI is expressed and cuts at one of the sites flanking the cassette in the cassette plasmid. The formation of a DSB recruits the homologous recombination machinery of yeast[90], which helps to find the homology on the target plasmid[91], leading to transfer of the cassette mutation into the target gene.

Romanini first demonstrated that the system could fix a non-sense stop codon in TRP1 with the wild-type Arg codon with a maximal 6% efficiency. This repair required the presence of both the I-SceI plasmid and the cassette flanked by I-SceI sites. He then showed that the cassette plasmid could be recovered from daughter cells intact, following successful recombination, indicating that enough intact copies survived DSB formation and could be transmitted to other yeast cells in later rounds. To showcase the mating element, Romanini moved to a more complicated experimental design and showed that stop codons in two separate genes on the same target plasmid could be fixed with two sequential

rounds of mating and induction of homologous recombination. This work proved the concept that mutational information on cassettes could be transferred horizontally, used to make mutations in a target gene, and then passed on again.

1.8 A model system for investigating affinity maturation

This heritable recombination system is well suited for the affinity maturation of binders. The structure of binding proteins is such that the flexible loops (complementarity determing regions or analogs) targeted for mutagenesis often form the key interactions with the ligand. Since the residues within the loops are located in consecutive regions of the primary sequence, a single cassette can randomize multiple key residues in a single mutational step. Additionally, an envisioned heritable recombination system for mixing and matching different binding loop sequences has a natural parallel in the combinatorial V(D)J system that is responsible for the expansive repertoire of the human adaptive immune system[92,93]. In T and B lymphocytes, the genes for the corresponding antigen receptors (immunoglobulin in B cells, T-cell receptors in T cells) are made functional when a VDJ recombinase enzyme recognizes recombinase activating segments that flank each of the V, D, and J cassettes in the germline DNA and catalyzes the joining of one of each of the three cassettes. In the Ig heavy chain, for example, 1 out of 100 V segments, 1 out of 27 D segments, and 1 out of 6 J segments come together, with 1.62×10^4 possible combinations; further combination with the light chain give a combinatorial diversity of 10^6 for a naïve Ig. Given the wide variety of antigens that can be bound by antibody with this system, it is reasonable to imitate this strategy of combinatorial construction of a binding interface.

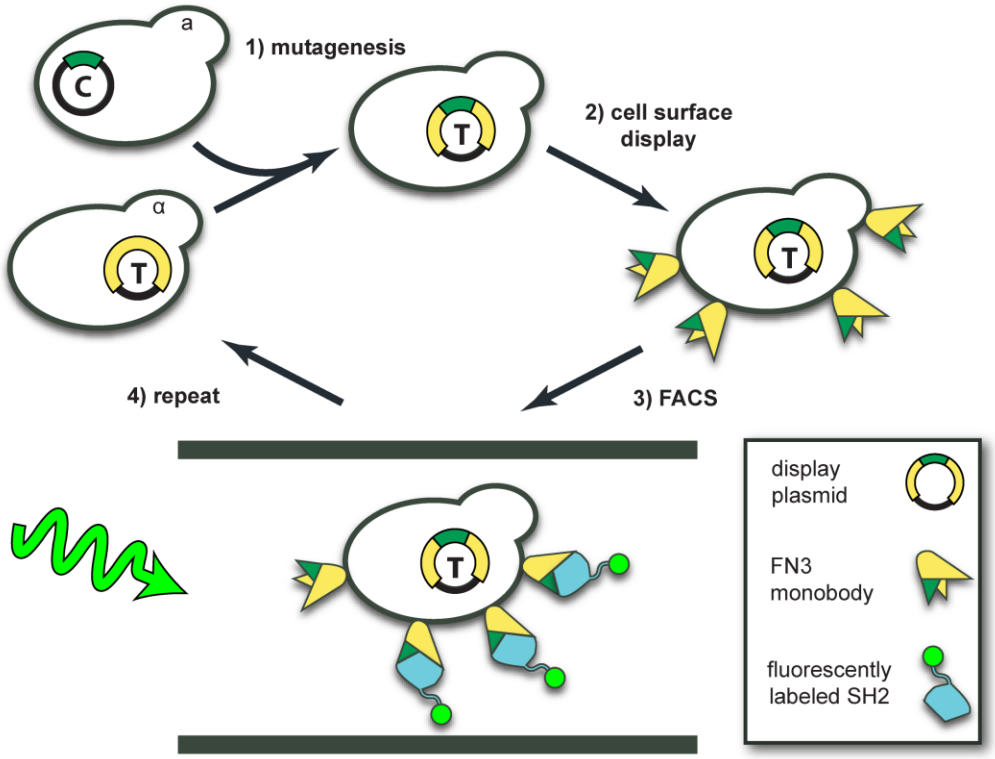


Figure 1-4. Heritable recombination system for affinity maturation of protein binders. Yeast mating brings together a monobody gene on a yeast display plasmid ("T" for target) and a cassette plasmid containing a mutagenic DNA sequence for one of the monobody loops. The cassette diversity replaces the corresponding loop in the monobody gene through homologous recombination, and then yeast display is induced, causing the new, mutated version of the monobody to be expressed on the cell surface. Selection (either magnetic beads or FACS is used to pick out the yeast cells with the highest affinity protein variants on their surface. Additional rounds of mating allow for repeated cycles of mutagenesis and selection.

In the following chapters, we discuss the construction of a heritable recombination system for affinity maturation of monobodies against the SH2 domain (Figure 1-4). This system combines a mutagenesis system, a selection system, and mating in yeast, allowing for fully *in vivo* cycles of affinity maturation. This system is a significant next step in that it demonstrates the applicability of HRS in more technically demanding situations. Instead of mutating a single codon to repair stop codons, this project requires the mutation of entire loops (8 or 9 codons) at a time. Furthermore, due to the nature of the

selection, it is necessary to maintain the target plasmid (containing the gene to be mutated) on a low copy CEN plasmid, instead of a high copy 2μ plasmid, and this comes with technical challenges as will be discussed in Chapter 3.

This work is also important from a synthetic biology perspective. The yeast display system and the heritable recombination system are two independent genetic systems for carrying out a biologically artificial function. Synthetic biology's future success depends on the ability to put together multiple exogenous, artificial genetic elements together and have them retain their function while having the cell retain its viability. The platonic ideal goal is to have sets of standard signaling elements (promoters, repressors, enzymes, kinases) that are robust enough to operate under a wide range of conditions with predictable behavior and that can interface with each other, somewhat like the resistors, capacitors, and inductors that make up electrical circuits. To date, efforts have been modest and have underscored the real-life difficulties inherent in getting messy, stochastic systems like cells to accept components that they have not evolved with. By bringing together two genetic systems (yeast display and heritable recombination), we are pushing the envelope in this regard. In overcoming technical hurdles for this project, we have obtained valuable insight into the design of sophisticated genetic systems in eukaryotes.

1.9 References

1. Ingber, D. E. Cellular tensegrity: defining new rules of biological design that govern the cytoskeleton. *Journal of Cell Science* **104**, 613–627 (1993).

2. Greenspan, N. S. Cohen's Conjecture, Howard's Hypothesis, and Ptashne's Ptruth: an exploration of the relationship between affinity and specificity. *Trends in Immunology* **31**, 138–143 (2010).

3. Padlan, E. A., Abergel, C. & Tipper, J. P. Identification of specificity-determining residues in antibodies. *The FASEB Journal* **9**, 133–139 (1995).

4. Schreiber, G. & Keating, A. E. Protein binding specificity versus promiscuity. *Current Opinion in Structural Biology* **21**, 50–61 (2011).

5. Erijman, A., Aizner, Y. & Shifman, J. M. Multispecific Recognition: Mechanism, Evolution, and Design. *Biochemistry* null–null (2011).at <http://dx.doi.org/10.1021/bi101563v>

6. Hudson, P. J. & Souriau, C. Engineered antibodies. *Nature medicine* **9**, 129–34 (2003).

7. *Current Protocols in Molecular Biology*. (John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2001).doi:10.1002/0471142727

8. Ross, M. H. & Pawlina, W. *Histology: a text and atlas: with correlated cell and molecular biology*. (Lippincott Williams & Wilkins: 2006).

9. Chan, A. C. & Carter, P. J. Therapeutic antibodies for autoimmunity and inflammation. *Nature Reviews Immunology* **10**, 301–316 (2010).

10. Beck, A., Wurch, T. & Corvaïa, N. Editorial: Therapeutic Antibodies and Derivatives: From the Bench to the Clinic. *Current Pharmaceutical Biotechnology* **9**, 421–422 (2008).

11. Sidhu, S. S. & Fellouse, F. A. Synthetic therapeutic antibodies. *Nat Chem Biol* **2**, 682–688 (2006).

12. Carter, P. J. Potent antibody therapeutics by design. *Nature reviews. Immunology* **6**, 343–57 (2006).

13. Beck, A., Wurch, T., Bailly, C. & Corvaia, N. Strategies and challenges for the next generation of therapeutic antibodies. *Nature Reviews Immunology* **10**, 345–352 (2010).

14. Miersch, S. & Sidhu, S. S. Synthetic antibodies: Concepts, potential and practical considerations. *Methods* **57**, 486–498 (2012).

15. Skerra, A. Alternative non-antibody scaffolds for molecular recognition. *Current Opinion in Biotechnology* **18**, 295–304 (2007).

16. Chames, P., Van Regenmortel, M., Weiss, E. & Baty, D. Therapeutic antibodies: successes, limitations and hopes for the future. *British Journal of Pharmacology* **157**, 220–233 (2009).

17. Gilbreth, R. N. & Koide, S. Structural insights for engineering binding proteins based on non-antibody scaffolds. *Current Opinion in Structural Biology* **22**, 413–420 (2012).

18. Frejd, F. Y. Novel alternative scaffolds and their potential use for tumor targeted radionuclide therapy. *Targeted Radionuclide Tumor Therapy* 89–116 (2008).at <http://www.springerlink.com/index/pu743xu665627035.pdf>

19. Nygren, P.-A. & Skerra, A. Binding proteins from alternative scaffolds. *Journal of Immunological Methods* **290**, 3–28 (2004).

20. Hosse, R. J., Rothe, A. & Power, B. E. A new generation of protein display scaffolds for molecular recognition. *Protein Science* **15**, 14–27 (2006).

21. Wu, T. Te & Kabat, E. A. An analysis of the sequences of the variable regions of bence jones proteins and myeloma light chains and their implications for antibody complementarity. *Journal of Experimental Medicine* **132**, 211–250 (1970).

22. Koide, A., Bailey, C. W., Huang, X. & Koide, S. The fibronectin type III domain as a scaffold for novel binding proteins. *Journal of Molecular Biology* **284**, 1141–1151 (1998).

23. Main, A. L., Harvey, T. S., Baron, M., Boyd, J. & Campbell, I. D. The three-dimensional structure of the tenth type III module of fibronectin: An insight into RGD-mediated interactions. *Cell* **71**, 671–678 (1992).

24. Koide, S., Koide, A. & Lipovšek, D. Chapter six - Target-Binding Proteins Based on the 10th Human Fibronectin Type III Domain (10Fn3). *Protein Engineering for Therapeutics, Part B* **Volume 503**, 135–156 (2012).

25. Dickinson, C. D. *et al.* Crystal structure of the tenth type III cell adhesion module of human fibronectin. *Journal of Molecular Biology* **236**, 1079–1092 (1994).

26. Branden, C. & Tooze, J. *Introduction to Protein Structure*. (Garland Science: 1999).

27. Olson, C. A. & Roberts, R. W. Design, expression, and stability of a diverse protein library based on the human fibronectin type III domain. *Protein Science* **16**, 476–484 (2007).

28. Plaxco, K. W., Spitzfaden, C., Campbell, I. D. & Dobson, C. M. Rapid refolding of a proline-rich all-beta-sheet fibronectin type III module . *Proceedings of the National Academy of Sciences of the United States of America* **93** , 10703–10706 (1996).

29. Parker, M. H. *et al.* Antibody mimics based on human fibronectin type three domain engineered for thermostability and high-affinity binding to vascular endothelial growth factor receptor two. *Protein Engineering Design & Selection* **18**, 435–444 (2005).

30. Cota, E., Hamill, S. J., Fowler, S. B. & Clarke, J. Two proteins with the same structure respond very differently to mutation: the role of plasticity in protein stability. *Journal of Molecular Biology* **302**, 713–725 (2000).

31. Batori, V., Koide, A. & Koide, S. Exploring the potential of the monobody scaffold: effects of loop elongation on the stability of a fibronectin type III domain. *Protein Engineering Design and Selection* **15**, 1015–1020 (2002).

32. Sweeney, C. J. *et al.* Phase I study of CT-322, first Adnectin protein therapeutic and potent inhibitor of VEGFR-2, in patients (pts) with advanced solid tumors (ST). *J Clin Oncol (Meeting Abstracts)* **26**, 3523– (2008).

33. Dineen, S. P. *et al.* The Adnectin CT-322 is a novel VEGF receptor 2 inhibitor that decreases tumor burden in an orthotopic mouse model of pancreatic cancer. *BMC Cancer* **8**, 352 (2008).

34. Mamluk, R. *et al.* Anti-tumor effect of CT-322 as an Adnectin inhibitor of vascular endothelial growth factor receptor-2. *mabs* **2**, 199–208 (2010).

35. Molckovsky, A. & Siu, L. First-in-class, first-in-human phase I results of targeted agents: Highlights of the 2008 American Society of Clinical Oncology meeting. *Journal of Hematology & Oncology* **1**, 20 (2008).

36. Wojcik, J. *et al.* A potent and highly specific FN3 monobody inhibitor of the Abl SH2 domain. *Nat Struct Mol Biol* **17**, 519–527 (2010).

37. Hackel, B. J., Neil, J. R., White, F. M. & Wittrup, K. D. Epidermal growth factor receptor downregulation by small heterodimeric binding proteins . *Protein Engineering Design and Selection* **25** , 47–57 (2012).

38. Getmanova, E. V *et al.* Antagonists to human and mouse vascular endothelial growth factor receptor 2 generated by directed protein evolution in vitro. *Chemistry & biology* **13**, 549–56 (2006).

39. Hackel, B. J., Kapila, A. & Dane Wittrup, K. Picomolar Affinity Fibronectin Domains Engineered Utilizing Loop Length Diversity, Recursive Mutagenesis, and Loop Shuffling. *Journal of Molecular Biology* **381**, 1238–1252 (2008).

40. Bradbury, A. R. M., Sidhu, S., Dubel, S. & McCafferty, J. Beyond natural antibodies: the power of in vitro display technologies. *Nature biotechnology* **29**, 245–254 (2011).

41. Schofield, D. *et al.* Application of phage display to high throughput antibody generation and characterization. *Genome Biology* **8**, R254 (2007).

42. Colby, D. W. *et al.* Engineering Antibody Affinity by Yeast Surface Display. *Methods in Enzymology* **Volume 388**, 348–358 (2004).

43. Rothe, A., Hosse, R. J. & Power, B. E. In vitro display technologies reveal novel biopharmaceutics. *FASEB Journal* **20**, 1599–1610 (2006).

44. Gai, S. A. & Wittrup, K. D. Yeast surface display for protein engineering and characterization. *Current Opinion in Structural Biology* **17**, 467–473 (2007).

45. Dufner, P., Jermutus, L. & Minter, R. R. Harnessing phage and ribosome display for antibody optimisation. *Trends in Biotechnology* **24**, 523–529 (2006).

46. Lipovsek, D. & Plückthun, A. In-vitro protein evolution by ribosome display and mRNA display. *Journal of Immunological Methods* **290**, 51–67 (2004).

47. Irving, R. A., Coia, G., Roberts, A., Nuttall, S. D. & Hudson, P. J. Ribosome display and affinity maturation: from antibodies to single V-domains and steps towards cancer therapeutics. *Journal of Immunological Methods* **248**, 31–45 (2001).

48. Boder, E. T. & Wittrup, K. D. Yeast surface display for screening combinatorial polypeptide libraries. *Nature biotechnology* **15**, 553–557 (1997).

49. Chao, G. *et al.* Isolating and engineering human antibodies using yeast surface display. *Nat. Protocols* **1**, 755–768 (2006).

50. Gera, N., Hussain, M. & Rao, B. M. Protein selection using yeast surface display. *Methods* (2012).at <http://www.sciencedirect.com/science/article/pii/S1046202312000709>

51. Boder, E. T. & Wittrup, K. D. Optimal Screening of Surface-Displayed Polypeptide Libraries. *Biotechnology Progress* **14**, 55–62 (1998).

52. Boder, E. T., Midelfort, K. S. & Wittrup, K. D. Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity. *Proceedings of the National Academy of Sciences, USA* **97**, 10701–10705 (2000).

53. Bowley, D. R., Labrijn, A. F., Zwick, M. B. & Burton, D. R. Antigen selection from an HIV-1 immune antibody library displayed on yeast yields many novel antibodies compared to selection from the same library displayed on phage. *Protein Engineering Design & Selection* **20**, 81–90 (2007).

54. Der, B. S. & Kuhlman, B. From Computational Design to a Protein That Binds. *Science* **332**, 801–802 (2011).

55. Ackerman, M. *et al.* Highly avid magnetic bead capture: An efficient selection method for de novo protein engineering utilizing yeast surface display. *Biotechnology Progress* **25**, 774–783 (2009).

56. Feldhaus, M. J. *et al.* Flow-cytometric isolation of human antibodies from a nonimmune Saccharomyces cerevisiae surface display library. *Nature biotechnology* **21**, 163–170 (2003).

57. Lin, H., Abida, W. M., Sauer, R. T. & Cornish, V. W. Dexamethasone-Methotrexate: An Efficient Chemical Inducer of Protein Dimerization In Vivo. *Journal of the American Chemical Society* **122**, 4247–4248 (2000).

58. Lin, H. & Cornish, V. W. Screening and Selection Methods for Large-Scale Analysis of Protein Function. *Angewandte Chemie International Edition* **41**, 4402–4425 (2002).

59. Lin, H., Tao, H. & Cornish, V. W. Directed Evolution of a Glycosynthase via Chemical Complementation. *Journal of the American Chemical Society* **126**, 15051–15059 (2004).

60. Peralta-Yahya, P., Carter, B. T., Lin, H., Tao, H. & Cornish, V. W. High-Throughput Selection for Cellulase Catalysts Using Chemical Complementation. *Journal of the American Chemical Society* **130**, 17446–17452 (2008).

61. Tao, H. & Cornish, V. W. Milestones in directed enzyme evolution. *Current Opinion in Chemical Biology* **6**, 858–864 (2002).

62. Baker, K. *et al.* Chemical complementation: A reaction-independent genetic assay for enzyme catalysis. *Proceedings of the National Academy of Sciences* **99** , 16537–16542 (2002).

63. Bosley, A. D. & Ostermeier, M. Mathematical expressions useful in the construction, description and evaluation of protein libraries. *Biomolecular Engineering* **22**, 57–61 (2005).

64. Firth, A. E. & Patrick, W. M. GLUE-IT and PEDEL-AA: new programmes for analyzing protein diversity in randomized libraries. *Nucleic Acids Research* **36**, W281–285 (2008).

65. Voigt, C. A., Mayo, S. L., Arnold, F. H. & Wang, Z. G. Computational method to reduce the search space for directed protein evolution. *Proceedings of the National Academy of Sciences, USA* **98**, 3778–3783 (2001).

66. Voigt, C. A., Mayo, S. L., Arnold, F. H. & Wang, Z.-G. Computationally focusing the directed evolution of proteins. *Journal of Cellular Biochemistry* **84**, 58–63 (2001).

67. Arnold, F. H. Fancy footwork in the sequence space shuffle. *Nature biotechnology* **24**, 328–330 (2006).

68. Wong, T. S., Roccatano, D. & Schwaneberg, U. Challenges of the genetic code for exploring sequence space in directed protein evolution. *Biocatalysis and Biotransformation* **25**, 229–241 (2007).

69. Armstrong, K. A. & Tidor, B. Computationally mapping sequence space to understand evolutionary protein engineering. *Biotechnology Progress* **24**, 62–73 (2008).

70. Reetz, M. T., Kahakeaw, D. & Lohmer, R. Addressing the Numbers Problem in Directed Evolution. *ChemBioChem* **9**, 1797–1804 (2008).

71. Kazlauskas, R. J. & Bornscheuer, U. T. Finding better protein engineering strategies. *Nat Chem Biol* **5**, 526–529 (2009).

72. Wong, T. S., Roccatano, D. & Schwaneberg, U. Steering directed protein evolution: strategies to manage combinatorial complexity of mutant libraries. *Environmental Microbiology* **9**, 2645–2659 (2007).

73. Muller, B. H. *et al.* In vitro affinity maturation of an anti-PSA antibody for prostate cancer diagnostic assay. *Journal of Molecular Biology* (2011).at <http://www.sciencedirect.com/science/article/pii/S0022283611011326>

74. Orcutt, K. D. *et al.* Engineering an antibody with picomolar affinity to DOTA chelates of multiple radionuclides for pretargeted radioimmunotherapy and imaging. *Nuclear Medicine and Biology* **In Press,** , (2010).

75. Chen, Y. *et al.* Selection and analysis of an optimized anti-VEGF antibody: crystal structure of an affinity-matured fab in complex with antigen. *Journal of Molecular Biology* **293**, 865–881 (1999).

76. Schaller, M. D. & Schaefer, E. M. Multiple stimuli induce tyrosine phosphorylation of the Crk-binding sites of paxillin. . *Biochemical Journal* **360**, 57–66 (2001).

77. Hoelbl, A. *et al.* Stat5 is indispensable for the maintenance of bcr/abl-positive leukaemia. *EMBO Molecular Medicine* **2**, 98–110 (2010).

78. Ye, D., Wolff, N., Li, L., Zhang, S. & Ilaria, R. L. STAT5 signaling is required for the efficient induction and maintenance of CML in mice. *Blood* **107** , 4917–4925 (2006).

79. Grebien, F. *et al.* Targeting the SH2-Kinase Interface in Bcr-Abl Inhibits Leukemogenesis. *Cell* **147**, 306–319 (2011).

80. Skorski, T. BCR-ABL1 Kinase: Hunting an Elusive Target with New Weapons. *Chemistry & Biology* **18**, 1352–1353 (2011).

81. Pawson, T., Gish, G. D. & Nash, P. SH2 domains, interaction modules and cellular wiring. *Trends in Cell Biology* **11**, 504–511 (2001).

82. Filippakopoulos, P., Müller, S. & Knapp, S. SH2 domains: modulators of nonreceptor tyrosine kinase activity. *Current Opinion in Structural Biology* **19**, 643–649 (2009).

83. Pawson, T. & Nash, P. Protein–protein interactions define specificity in signal transduction. *Genes & Development* **14**, 1027–1047 (2000).

84. Gan, W. & Roux, B. Binding specificity of SH2 domains: Insight from free energy simulations. *Proteins: Structure, Function, and Bioinformatics* **74**, 996–1007 (2009).

85. Huang, H. *et al.* Defining the Specificity Space of the Human Src Homology 2 Domain. *Molecular & Cellular Proteomics* **7**, 768–784 (2008).

86. Ladbury, J. E. & Arold, S. Searching for specificity in SH domains. *Chemistry & Biology* **7**, R3–R8 (2000).

87. Pershad, K. *et al.* Generating a panel of highly specific antibodies to 20 human SH2 domains by phage display. *Protein Engineering Design & Selection* **23**, 279–288 (2010).

88. Yang, W.-P. *et al.* CDR Walking Mutagenesis for the Affinity Maturation of a Potent Human Anti-HIV-1 Antibody into the Picomolar Range. *Journal of Molecular Biology* **254**, 392–403 (1995).

89. Romanini, D. W., Peralta-Yahya, P., Mondol, V. & Cornish, V. W. A Heritable Recombination System for Synthetic Darwinian Evolution in Yeast. *ACS Synthetic Biology* 121024121138004 (2012).doi:10.1021/sb3000904

90. Krogh, B. O. & Symington, L. S. RECOMBINATION PROTEINS IN YEAST. *Annual Review of Genetics* **38**, 233–271 (2004).

91. Barzel, A. & Kupiec, M. Finding a match: how do homologous sequences get together for recombination? *Nature reviews. Genetics* **9**, 27–37 (2008).

92. Abbas, A. K., MD, A. H. L. & Pillai, S. *Cellular and Molecular Immunology*. 576 (Saunders: 2009).at <http://www.amazon.com/Cellular-Molecular-Immunology-Updated-STUDENT/dp/1416031235>

93. Glanville, J. *et al.* Precise determination of the diversity of a combinatorial antibody library gives insight into the human immunoglobulin repertoire . *Proceedings of the National Academy of Sciences* **106** , 20216–20221 (2009).

Chapter 2

Cconstruction of BC loop libraries and Phenotypic Selection

2.1 Abstract

One half of the directed evolution equation is selection of functional binders. In order to develop a system for mutagenesis, we first needed to validate the components of the functional selection. We started by purifying our target GST-SH2 and functionalizing it with biotin. We then demonstrated that magnetic beads coated with our target could enrich our high-affinity monobody from a mock library. The high-affinity phenotype was also successfully screened for by fluorescent labeling and flow cytometry. To test these selections in an actual round of affinity maturation, we mutated the BC loop of a medium-affinity binder using gap-repair and showed using FACS that rare hits initially present at <0.055% in a 6.5 x 10^6 library could be enriched. In short, we demonstrated that all of the components of the selection system were working, in addition to validating our choices of library diversity. The data gave us sufficient confidence that this selection system could be relied upon for troubleshooting and development of a novel method for *in vivo* mutagenesis.

2.2 Introduction

The heritable recombination system for affinity maturation, as discussed in Chapter 1, is a sophisticated genetic system, containing a separate mutagenesis system and a selection system. At the beginning of this project, we first wanted to test the components of the selection system before trying to interface it with the mutagenesis system. Specifically, we needed to test the strain and plasmid required for yeast display on the cell surface, express and purify the Abl1 SH2 target, demonstrate the ability of both the magnetic beads and FACS functional screens to enrich binders from libraries, and show that we could build and test BC loop library mutants with all of these put together. To provide a short-term goal to motivate these efforts, we decided to try to affinity mature a medium-affinity monobody. If we could successfully accomplish all of these steps, then we could have confidence that the selection was working

robustly. Since the Cornish Lab had no previous experience with yeast display, magnetic beads, or FACS, all of the materials needed to be procured and the protocols adapted and optimized.

The yeast display system, developed by Boder and Wittrup[1], physically links genotype and phenotype in yeast cells, allowing for phenotypic selection of libraries of genes. There are two components of this system: the EBY100 strain of yeast and the pYD-1 plasmid. A monobody gene can be inserted into pYD-1 as an in-frame fusion with the AGA2 gene. When pYD-1 is transformed into EBY100, growth of the resulting strain in galactose media induces expression of the monobody protein, fused to aga2p on the surface of the yeast.

Another key component needed for the project was the gene for the monobody protein scaffold. Our collaborators in the Koide Lab at the University of Chicago sent us the gene for the high-affinity HA4 binder of Abl1 SH2[2], in addition to a lower affinity variant (Kd ≈ 350 nM) of HA4 called "HA4R38A". These monobodies served as controls in the work described below. They also provided us with the plasmid for expressing GST-SH2, which we need to purify and biotinylate in order to for the phenotypic selections to work.

The magnetic bead selection is a critical technique for processing new libraries, and has been well characterized by Ackerman and colleagues[3]. It is high throughput, with the ability to run 10^9 yeast cells in a single 1.5 mL microfuge tube. This means that after a recent round of mutagenesis, when the library complexity is highest but many non-functional mutants are present, magnetic bead selection can quickly and easily cull the non-binders from the population before running the more expensive and (relatively) low-throughput FACS. An additional feature of the magnetic bead technique is that it utilizes avidity, as a single bead is coated with many copies of antigen and a single yeast cell is coated with many copies of binder. This means that the selection can even enrich binders with moderate function. This is important for an affinity maturation scheme where multiple steps of mutation are envisioned (such as ours), and where weaker binders might actually be critical intermediates on the mutational pathway to very strong binders, and so we would prefer not to weed them out immediately. That said, stringency can be tuned upwards if desired by increasing the number of wash steps, the temperature of the wash buffer, or the speed of mixing. Ackerman also demonstrated that beads could retain rare clones. She did

this by mixing a micromolar (low affinity) binder with non-binding cells in ratios as rare as 1 in 10^5, and then showing that the binding clone was enriched to a post-selection prevalence of 1 in 3.

Finally, FACS would ultimately be needed in order to identify the very high affinity binders. FACS is well-suited for fine affinity discrimination. Importantly, it allows researchers to control for the copy-number variation of binders on the surface when analyzing the antigen binding signal. It has the added benefit that it can be used to make an affinity determination for the a population of cells expressing the same clone[4,5].

Using magnetic beads and FACS in combination is key. There are stochastic elements in the FACS process that could cause rare cells to be lost, even if they have very high-affinity binders. For instance, sometimes two cells sticking to each other will go through the detection laser at the same time. Forward scattering and back scattering measurements by the instrument will be used to automatically discard these cells since they don't appear to have the shape and size of the rest of the yeast that are being examined. It is critical, then, to perform a low-stringency selection to increase the multiplicity of the cells possessing functional binders, prior to running FACS.

2.3 Results

Protein Expression

To start testing the functionality of the selections, we first needed to purify the SH2 domain that the monobodies targeted. From the Koide Lab, we obtained plasmid pGST-SH2 for expressing a fusion protein of the human Abl1 SH2 domain and glutathione-S-transferase (GST)[2]. Fusion to GST increases the solubility of the protein and provides a purification alternative to the N-terminal 6-histidine tag. The GST domain is located on a face of SH2 opposite from the HA4 binding epitope, so it does not interfere with the HA4 interaction. The GST domain also contains an N-terminal cysteine residue for biotin-tagging chemistry. We expressed the GST-SH2 fusion protein in BL21 DE3 *E. coli* cells using the auto-induction protocol of Studier[6] (Figure 2-1a). The GST-SH2 fusion was purified by its 6-histidine tag on a Ni^{2+} Sepharose ion affinity column (Figure 2-1b). On average, each 500 mL of auto-induction culture yielded

4-5 mg of GST-SH2 after a single column purification. More GST-SH2 was expressed than this however, and multiple column purifications of "dirty" fractions could be used to enhance yield.

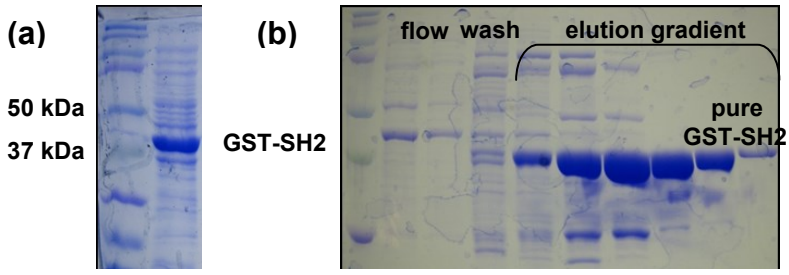


Figure 2-1. Expression and purification of GST-SH2 fusion protein. **a,** GST-SH2 is visibly over-expressed after auto-induction overnight. **b,** FPLC elution gradient is capable of producing pure GST-SH2.

The free cysteine in the GST domain was utilized to covalently attach biotin to the GST-SH2. In both the magnetic bead selection and FACS screening, the biotin is used as a handle for the SH2 domain. The magnetic beads are coated with streptavidin (which binds biotin with $K_d = 4x10^{-14}$ M)[7], and in FACS, fluorescent phycoerythrin is conjugated to streptavidin. In this way, SH2 can be attached to a solid-phase or fluorescently labeled, along with any yeast that happen to be binding it.

Originally, we used a biotinylation reagent with maleimide moiety for the covalent linkage. After reaction, we used a HABA (4'-hydroxyazobenzene-2-carboxylic acid) colorimetric assay to estimate the molar ratio of biotin:GST-SH2. After multiple trials of reactions at different conditions, we obtained molar ratios ranging from 1.7 – 4.3. There is only one cysteine in the translated GST-SH2 protein, so we expected a molar ratio of about 1:1. Higher ratios imply that the average GST-SH2 molecule is biotinylated at more than one residue. Since the maleimide moiety in the biotinylation reagent is potentially reactive towards lysine residues as well, over-biotinylation might signify the presence of unwanted biotins at the HA4 binding epitope that could ruin the selection. As a result, I switched to a disulfide biotinylation reagent, which is more specific for the cysteine. Though this conjugation is reversible by reduction, we anticipated that this would not be a problem for our applications.

Magnetic Bead Selections

With biotinylated GST-SH2 in hand, we wanted to test both the yeast display of HA4 and the magnetic bead selection. We cloned the high-affinity HA4 monobody into the yeast display plasmid which was then transformed into EBY100 yeast to make ysEBYHA4. For a negative control, we made a strain displaying a non-sense truncation monobody (ysEBYR38A*), and for a positive control we made a strain (ysEBYL751) displaying an extremely high-affinity monobody developed by the Wittrup Lab that targets lysozyme[8]. The first experiment examined whether cells could display monobody and stick to the magnetic beads (Table 2-1). We induced yeast display in each of the three cell strains and incubated them separately with magnetic beads. After the initial incubation of the cells with the beads and magnetic pull-down, we removed the supernatant (cells not stuck to beads) and dilution plated it in order to quantify cell count. After washing the beads four times to remove non-specific binders, we dilution plated the captured cell population.

Table 2-1. *Magnetic bead binding capacities.*

Strain	Kd	Cells in Supernatant	Cells Captured	Beads/Cell
ysEBYHA4	7 nM	2×10^7	2.7×10^6	1.9
ysEBYL751	3 pM	1×10^7	1×10^6	5
ysEBYR38A*	n.d.	5.3×10^8	1.4×10^4	360

$5x10^6$ target-coated magnetic beads were mixed with $5x10^8$ cells of each of the three strains displaying a different antibody. “Cells in supernatant” is the number of cells not bound to the beads after 2.5 hrs incubation. “Cells captured” is the number of cells attached to beads after 4x washes. Significant figures based on number of colonies on dilution plates.

We found that for both the positive control and the HA4-expressing yeast cells, the beads capture a total number of cells with the same order of magnitude (10^6). Further experiments have supported the finding that 2:1 beads:cell is the saturation point of the beads. It may seem at first glance surprising that lower-affinity HA4 stuck in larger total numbers than L7.5.1, but consideration of binding avidity between cells and beads[3] means that any reasonably high-affinity binders should be expected to perform equally

well. The difference in performance seen here is within experimental error. Some negative control cells still stuck to the beads after 4 washes. Presumably under conditions where there is no competition, the stickiness of yeast allows this background level of selection. In the bead selections using actual libraries discussed later in this and following chapters, libraries were first incubated with uncoated beads as a negative selection for sticky mutants; unbound cells were then transferred to a tube with SH2-coated beads, whereas the cells that had bound uncoated beads were discarded. The much higher performance of ysEBYHA4 vs ysEBYR38A* suggests that the yeast display, the biotinylated SH2, and the magnetic beads were all working as anticipated.

Before running real libraries, we needed to check that we could successfully enrich rare binders from a background of non-binding cells with my experimental setup. Thus, the next step was to test the ability of the beads to enrich high-affinity binders from a "mock library" of EBY100 cells containing no display plasmid. 5×10^8 yeast cells, consisting of a mixture of ~1000 background cells for every monobody-displaying cell, were selected with 2.5×10^6 antigen-coated beads. The pre- and post-selection frequencies of the monobody-displaying cells was determined by dilution plating on selective media.

Table 2-2. *Magnetic bead enrichments of mock libraries in cell frequencies.*

Monobody	Pre Fraction (%)	Post Fraction (%)	Enrichment Ratio	Theoretical Max Ratio
HA4	0.13	93.1	695	746
L7.5.1	0.14	61.6	440	710
HA4R38A*	0.17	<0.14	-	590

$5x10^5$ yeast display cells with expressing a specific monobody were mixed with $5x10^8$ EBY100 cells. Selections were carried out with $2.5x10^6$ beads. Fractions (# plasmid-containing cells / # all cells) were determined both pre- and post-selection by dilution plating on SC(U-) and SC(UT-). Enrichment ratio is post fraction / pre fraction. Theoretical max ratio is the enrichment ratio that would be required to enrich the monobody-displaying cells from the observed pre fraction to a post fraction of 100% (purity).

Table 2-3. *Magnetic bead enrichments of mock libraries in absolute numbers*

Monobody	Total Pre	Monobody Pre	Total Post	Monobody Post
HA4	9.02×10^8	1.21×10^6	2.16×10^5	2.01×10^5
L7.5.1	1.1×10^9	1.52×10^6	4.6×10^5	2.82×10^5
HA4R38A*	1.1×10^9	1.86×10^6	7×10^3	< 10

The same experiment from Table 2-2, with data in terms of absolute cell numbers. "Total" is the total number of yeast (background + monobody-displaying).

The results show that both ysEBYHA4 and ysEBYL751 enriched to a majority of the post-selection population (Table 2-2), successfully competing with the non-binding yeast. Both ysEBYHA4 and ysEBYL751 appear to have reached the level of bead saturation, with about 2:1 beads:monobody cells captured. ysEBYL751 performed better than ysEBYHA4 in number of monobody cells pulled down, but it also apparently pulled a lot of non-binding cells too, resulting in a lower post-enrichment frequency than ysEBYHA4 .The negative control, expressing non-functional truncated protein was not enriched at all. No ysEBYR38A* colonies were found on post-selection plates, so post-enrichment fraction in Table 2-2 and monobody post in Table 3-3 are upper-bounds. These results importantly showed that the magnetic bead selection worked well even with an input of 10^9 cells in a single microfuge tube, and is thus very useful as a high-throughput initial selection on a large library.

Flow cytometry of HA4

The other selection technology that this project relies upon is flow cytometry. We validated the use of flow cytometry by testing it on strain ysEBYHA4 (Figure 2-2). We induced yeast display in a culture of ysEBYHA4 and then incubated 10^6 cells with biotinylated GST-SH2. Our yeast display construct contains a V5 epitope tag C-terminal to the monobody, and we used an anti-V5 antibody conjugated to fluorescein isothiocyanate (FITC) to fluorescently label this tag. Doing this allows us to use the FITC signal as a measure of how many copies of the monobody fusion are being displayed on the surface of the yeast. Fluorescent phycoerythrin(PE)-streptavidin was used to label the SH2 domain, and thus the PE signal measures the amount of ligand binding to the cells.

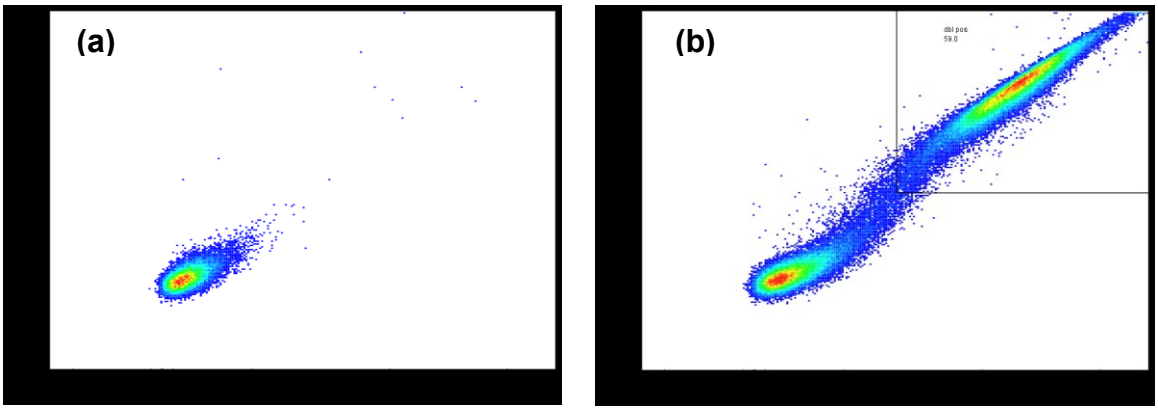


Figure 2-2. HA4 monobodies expressed on the surface of yeast give double-positive signal in flow cytometry. FITC signal for anti-V5 antibody is plotted in log units on the X-axis. PE signal for GST-SH2 is plotted in log units on the Y-axis **a,** EBY100 negative control cells; all cells are double-negative. **b,** 59% of ysEBYHA4 cells show a double-positive signal. Heat map colors represent data-point density, with red > green > blue.

The results showed that anti-V5 antibody binds to galactose-induced ysEBYHA4, confirming the expression of the Aga2p-monobody fusion on the surface of the yeast. Additionally, the positive PE signal suggests that biotinylated GST-SH2 can bind to both the HA4 monobody and to streptavidin. As we would expect, binding to SH2 is proportional to the amount of monobody expressed on the surface of the yeast. Furthermore, a lack of PE single-positive events suggests that there is no non-specific binding of our GST-SH2 to the yeast cell. It is typical of yeast display that only 40-80% of cells capable of yeast display will actually express binder on the surface after galactose induction[9], so the 59% of ysEBYHA4 giving double-positive signal is an expected result. With the two selections and the yeast display system working, we were at this point ready to start building and sorting libraries.

Construction of BC loop libraries

We then began construction of a few monobody libraries using established cassette mutagenesis methods[8,10] so that we could test the selection methods in a real application. The clone that we chose to affinity mature was HA4R38A with $K_d \approx 350$ nM.

The BC loop of the monobody structure looked appropriate target for cassette mutagenesis. Cassette mutagenesis is ideal for randomizing a consecutive set of residues, and so the β-hairpin loops are natural functional units to target. The crystal structure for the HA4 clone and alanine scanning data obtained by Wojcik indicate that the FG loop makes more energetic contribution to binding[2]. Because of this, we reasoned that complete replacement of the FG loop would really not be affinity maturation at all, but more like de novo binder discovery due to the removal of key binding contacts. The crystal structure also shows that only Tyr35 from the BC loop makes direct contact with the SH2 domain. However Ser32-Tyr35 of the BC loop directly contact the FG loop. It is possible then that mutation of the BC loop could modulate the interaction of the FG loop with the SH2 domain. For example, a mutant BC loop might favorably change the entropy change of binding by holding the FG more closely to an optimal geometry.

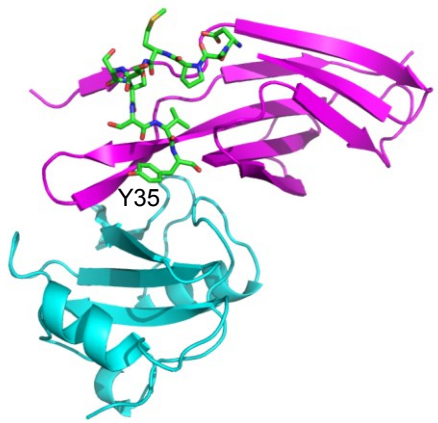


Figure 2-3. BC loop residues targeted for mutagenesis.
The HA4 protein (magenta) is shown in complex with the Abl1 SH2 domain (cyan). The BC loop, with residue sequence DAPMSSSSVY, is represented explicitly by sticks. The Tyr35 residue makes close contacts with the SH2 domain.
Image made with PDB 3K2M in PyMOL.

We built our first library by replacing the BC DNA sequence with 8 randomized codons. Oligonucleotides were synthesized using NNB degenerate nucleotides, where N indicates an equiprobable mix of the four nucleotides and B indicates C, G, or T. Such a degenerate codon has the ability to code for all 20 amino acids, with only a 2% chance of a stop codon. The yeast display plasmid was restriction digested and cotransformed into yeast with the synthetic oligo library (Figure 2-4).

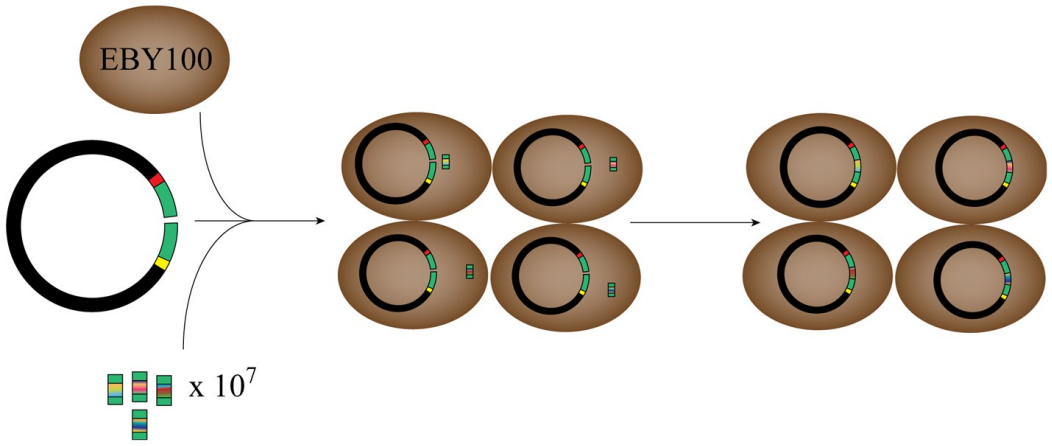


Figure 2-4. Gap-repair mutagenesis for construction of BC loop libraries. Yeast display plasmid (black ring) was restriction digested in the center of the monobody gene (green). Mutagenic cassettes were constructed by PCR with randomized codons, giving each cassette a unique sequence (rainbow patterns) flanked on both sides by homology to the monobody gene (green). The cut plasmid and the cassettes were co-transformed into EBY100 yeast, with gap-repair by homologous recombination resulting in the desired yeast library.

2.0×10^7 successful transformants were obtained, and this serves as an approximation to the true library complexity. It should be noted that this actual library complexity vastly undersamples the potential library complexity of 8 NNB codons. With 48 codon possibilities at the nucleotide level, we calculate $48^8 = 2.8 \times 10^{13}$ unique nucleotide sequences. For every sequence that we have in the library, there are a million that are missing from the potential library. Furthermore, each of the eight NNB codons has a 2.08% of coding for a stop. This means that 15.5% (calculated as $1 - 0.9792^8$) of the potential library members are expected to be useless since they will contain at least 1 or more stop codons in the BC loop.

We then sequenced a random sample of the transformed clones to assess the quality of the library. PCR from 40 colonies resulted in 20 quality (high signal to noise) sequencing reads (Table 2-4).

Clone	BC Loop Position							
FN3	D	A	P	A	V	T	V	R
HA4	D	A	P	M	S	S	S	S
1	G	P	E	N	R	V	N	A
2	G	G	P	T	D	V	D	T
3	T	V	L	L	P	I	L	V
4	T	R	*	A	*	P	F	Q
5	W	G	R	P	R	N	P	R
6	S	P	L	R	P	L	T	S
7	R	V	S	D	R	V	N	P
8	A	S	A	A	D	L	T	G
9	P	A	P	V	R	L	R	R
10	P	N	L	P	G	P	L	P
11	G	T	R	P	I	H	H	E
12	W	G	R	H	R	N	P	R
13	D	G	P	F	S	R	P	A
14	P	T	P	A	G	Y	A	F
15	H	W	L	K	P	H	R	N
16	L	T	L	V	Y	R	M	I
17	D	V	L	G	P	N	A	S
18	D	D	T	L	H	R	R	P
19	G	Q	A	S	H	C	G	E
20	P	S	A	P	P	P	A	P

Table 2-4. BC residues sequence from colony PCR of 8x NNB library.

BC residues sequences are shown for the wild-type fibronectin scaffold,

HA4 clone, and 20 members of the NNB codon library. * = stop codon.

On visual inspection, the large number of proline and arginine residues stood out. 17/20 had at least one proline, and 7/20 had 2 or more. The codon frequencies for these 20 sequences were calculated and compared to the theoretical frequencies for an NNB library (Table 2-5).

Table 2-5. Observed and expected residue frequencies in NNB library

Residue	# Observed	Observed (%)	Expected Freq (%)	Δ%
Ser	8	5	10.42	-5.42
Cys	1	0.625	4.17	-3.55
Tyr	2	1.25	4.17	-2.92
Phe	3	1.875	4.17	-2.3
Ile	3	1.875	4.17	-2.3
Lys	1	0.625	2.08	-1.46
Met	1	0.625	2.08	-1.46
Gln	2	1.25	2.08	-0.83
Stop	2	1.25	2.08	-0.83
Val	9	5.625	6.25	-0.63
Glu	3	1.875	2.08	-0.21
Trp	3	1.875	2.08	-0.21
Thr	10	6.25	6.25	0
His	7	4.375	4.17	0.21
Leu	14	8.75	8.33	0.42
Asp	8	5	4.17	0.83
Asn	8	5	4.17	0.83
Gly	13	8.125	6.25	1.88
Ala	13	8.125	6.25	1.88
Arg	20	12.5	8.33	4.17
Pro	29	18.125	6.25	11.88
Total	160	100%	100%	0

The 8 codons from the 20 sequences were tabulated, and frequencies were calculated.

Δ% = observed % - expected %

This breakdown shows greater than expected frequencies of Pro and Arg residues, and lower than expected frequencies of Ser, Cys, and Tyr. Although proline residues can often be found in β-hairpin turns, an abundance of prolines are expected to be conformationally restrictive. A lower frequency of Ser and Tyr is undesirable, as these residues tend to be enriched at binding interfaces[11–13]. Cys residues

have been found to make substantial monobody libraries[14], but it is not clear how general this result is and whether Cys is to be desired in our libraries. Overall, the observed distribution seems unfavorable for a binding library.

A χ^2 test rejected the hypothesis that the expected NNB residue frequency describes our library with $p < 0.005$ ($\chi^2 = 68.8$ with 17 degrees of freedom). Although this library incoporated all 20 amino acids as intended, the frequencies for some important amino acids were heavily skewed. This bias could have been introduced at multiple points in the process of building the library: during oligo synthesis, cassette PCR or gap-repair.

For a second library, we restricted the amino acid diversity to a subset of the natural 20 in order to decrease the total theoretical complexity. Our aim was to make a "smarter" designed library, where each individual library member would have a higher *a priori* probability of coding for a high-affinity binder[15,16]. This strategy has been proven effective even in the extreme case where binding loop libraries are restricted to just serines and tyrosines, so-called binary interfaces[11,12,17]. There are many options available for creating restricted diversity libraries, and construction using degenerate codons is greatly aided by computational tools that have been published and provided as an internet resource[18,19]. One way to improve the quality of the library is to completely remove stop codons; for our 8 codon BC loop, 16% of usable library size could be reclaimed by switching from NNB to a stop-free codon. It is also desirable, for the reasons stated above, to have libraries enriched in Tyr and Ser residues.

We found the NNY codon to be particularly attractive. The residue frequencies for an NNY library were favorable compared to the commonly used NNB and NNK codons. Importantly, there were no stop codons, so truncations would not be taking a 16% bite out of our usable library like in the NNB case. Furthermore, Tyr has a 50% increased frequency and Ser has a 20% increased frequency in NNY over NNB. Glu, Lys, Trp, Gln, and Met are absent, but they have counterparts with similar physicochemical properties (suggested by the residue groupings in Table 2-6) that are still represented. Thus I made an 8 residue NNY library in the same manner as the NNB library. Transformation into yeast gave 1.9×10^6 transformants.

Table 2-6. Degenerate codons and their corresponding residue frequencies

	NNN	NNK	NNB	NNY
Aspartic Acid	3.125	3.13	4.17	6.25
Glutamic Acid	3.125	3.13	2.08	0
Arginine	9.375	9.38	8.33	6.25
Histidine	3.125	3.13	4.17	6.25
Lysine	3.125	3.13	2.08	0
Phenylalanine	3.125	3.13	4.17	6.25
Tyrosine	3.125	3.13	4.17	6.25
Tryptophan	1.5625	3.13	2.08	0
Serine	9.375	9.38	10.42	12.5
Threonine	6.25	6.25	6.25	6.25
Cysteine	3.125	3.13	4.17	6.25
Asparagine	3.125	3.13	4.17	6.25
Glutamine	3.125	3.13	2.08	0
Glycine	6.25	6.25	6.25	6.25
Alanine	6.25	6.25	6.25	6.25
Proline	6.25	6.25	6.25	6.25
Valine	6.25	6.25	6.25	6.25
Leucine	9.375	9.38	8.33	6.25
Isoleucine	4.6875	3.13	4.17	6.25
Methionine	1.5625	3.13	2.08	0
Amber	1.5625	3.13	2.08	0
Ochre	1.5625	0.00	0.00	0
Opal	1.5625	0.00	0.00	0

The length of the mutated loops has also been an important source of diversity for the generation of high-affinity binders[8,20]. For this reason, we also constructed NNB and NNY libraries in which 4 loop lengths (6, 7, 8, or 9 codons) were present in an equimolar mixture. Hackel et al. used these 4 BC loop lengths to find a 3 pM binder against lysozyme[8]. He investigated 16 different fibronectin type III domains across 5 different species and found that these were the most common BC loop lengths. In the construction of our mixed-length libraries, I used an equimolar mixture of the 4 lengths not knowing *a*

priori which length would be the best design. The NNB mixed-length library had $3.8x10^6$ transformants, and the NNY mixed-length library had $6.5x10^6$ transformants.

To confirm proper construction of these two libraries, I extracted whole population DNA via PCR and then sequenced the monobody gene. Example data for the 8-codon NNB library is shown in Figure 2-5a for comparison. The 24 bp stretch of NNB and NNY nucleotides was confirmed for both libraries. Although peak heights cannot be used to quantitatively compare molar proportions of the different nucleotides, a qualitative excess of cytosine bases, as suggested in the data for both libraries, would be consistent with the observed overabundance of proline residues (CCN codons) noted above in the sequencing data for individual clones of the NNB library.

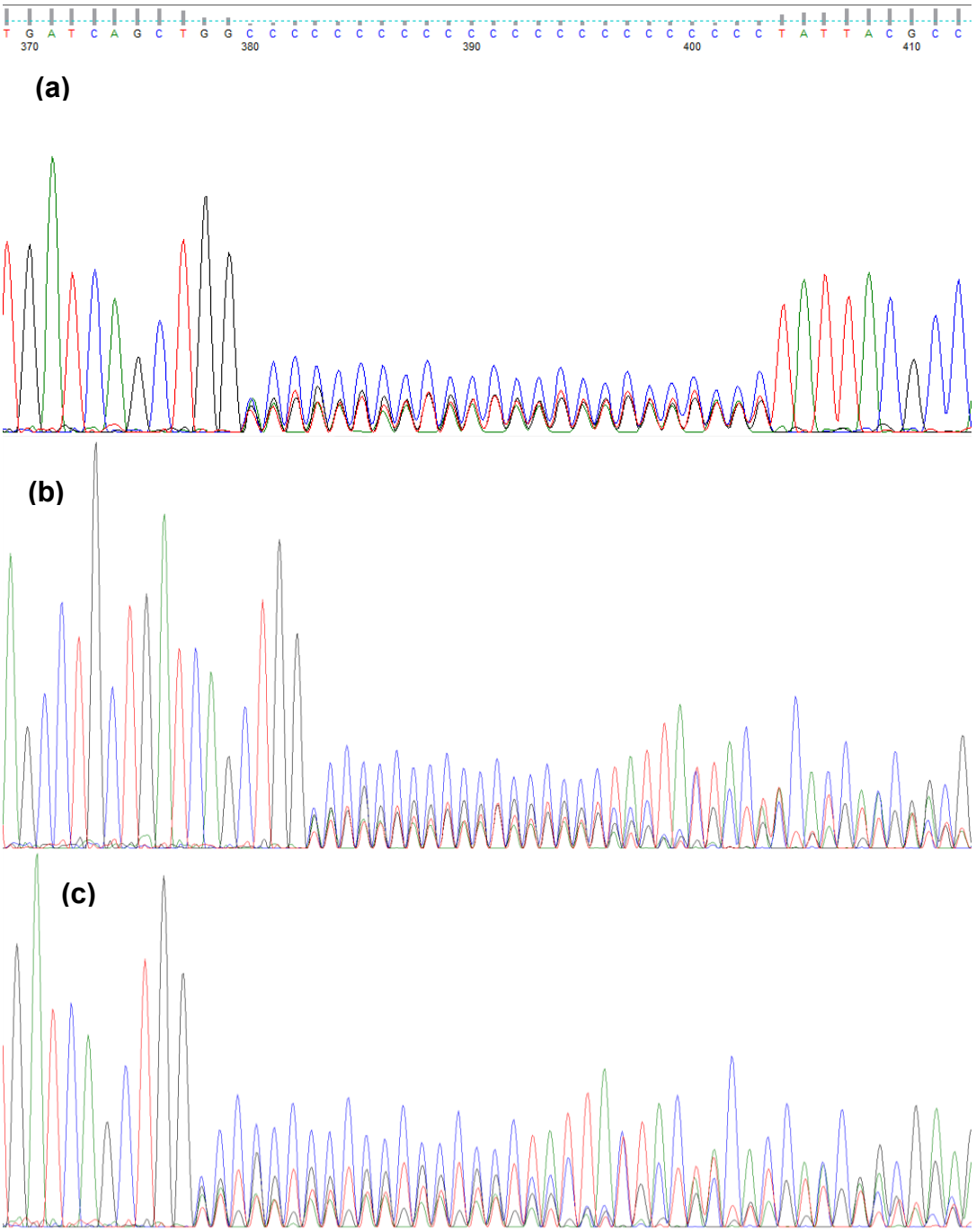


Figure 2-5. Population sequencing of BC loop coding regions. **a,** 8-codon NNB library, where green signal for A is flat every third position. **b,** mixed-length NNB library. **c,** mixed-length NNY library, note presence of G (black) signal every third base, contrary to design. In the mixed-length sequence reads, the difference in length of random codons results in loss of sequence matching downstream of the BC region. Adenine signal in green, thymine in red, cytosine in blue, guanine in black. 'N' stands for A,C,T or G by design; 'B' = C,G, or T; 'Y' = C or T.

FACS Enrichment on Mixed-Length Libraries

We next used FACS screening on the two mixed-length libraries. The mixed-length NNB library enriched only minimally after two rounds of FACS (Figure 2-6). A small cluster of FITC negative, PE positive cells was observed (Figure 2-6b).

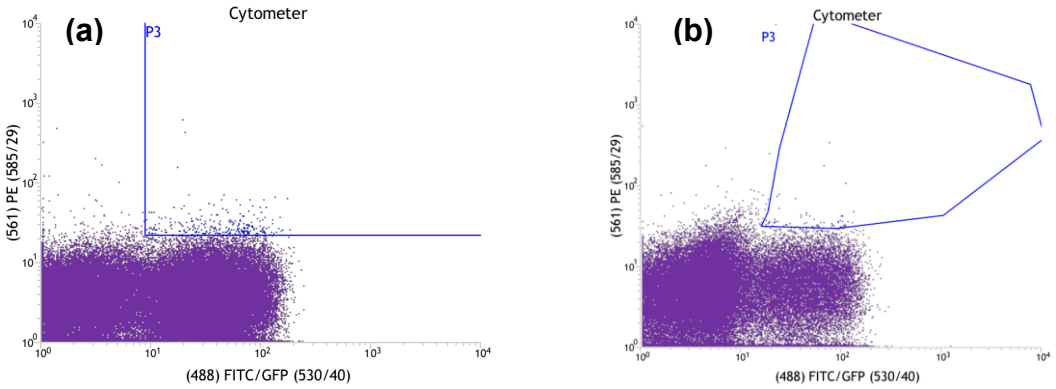


Figure 2-6. FACS enrichment of mixed-length NNB BC loop library. **a,** first round of sorting. 0.13% of cells were collected. **b,** second round of sorting. A small promontory of cells with PE-positive signal is seen just to the left of the blue P3 gate.

The mixed-length NNY library successfully enriched after three rounds (Figure 2-7). There is a distinct population of double-positive cells.

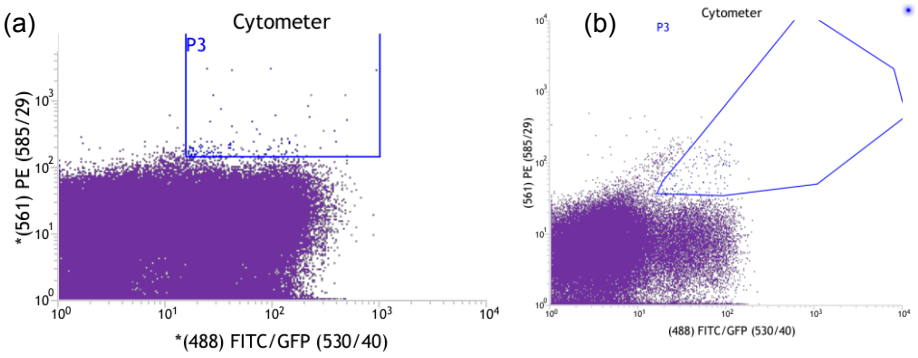


Figure 2-7. FACS enrichment of mixed-length NNY BC loop library. **a,** first round of sorting. (0.055%). **b,** third round of sorting. A population of double-positive hits is visible. 0.14% of cells collected.

2.4 Discussion

In this work, we found that yeast display is a very robust technology for working with libraries of protein binders. Magnetic beads were capable of capturing as many high-affinity cells as they could physically hold, even at extremely high cell densities (10^9 cells per tube). FACS was able to distinguish high-affinity binders from background and to enrich binding populations from rarity hits in actual BC loop libraries. Finally, the successful enrichment validated NNY codons as a source of diversity for affinity maturation. The fact that we obtained enrichment after loop replacement is validation of the NNY library as source of diversity for affinity maturation.

Our main goal at this early stage was to successfully build libraries with conventional gap-repair mutagenesis and then demonstrate enrichment. The affinity of the enriched binder populations from the mixed-length libraries could undoubtedly be improved further by many additional rounds of mutagenesis, involving at each step 1) extraction of the plasmid DNA from these yeast, 2) *in vitro* mutagenesis either by error-prone PCR or by chain-shuffling, and 3)transformation back into yeast[8]. However, our ultimate aim was to develop a new *in vivo* method for affinity maturation, and so we moved on after this initial proof of function to the development of the *in vivo* mutagenesis system described in Chapter 3.

The R38A mutant might be an interesting model system in the future for exploring evolution pathway divergence. We choose to build our first loop libraries with the medium-affinity R38A clone simply because we needed some libraries to test and hypothesized that for a single round of mutation, it might be easier to obtain an improvement in a medium-affinity binder than in a high-affinity binder. Although Arg38 contributes significantly ($\Delta\Delta G$ = 2.2 kcal/mol) to the binding of HA4, it is "merely" a scaffold residue. It was not mutated as a part of Wojcik's library to produce HA4 since it is located outside of the solvent-exposed BC and FG loops that are typically targeted in monobody evolution experiments. Rather, it is located within a β-strand in the framework region of the scaffold. It turns out that, partly because of the Arg38 interaction with SH2, the HA4 monobody binds SH2 in an atypical way, with the SH2 domain pinched between the FG loop and the framework CDFG β-sheet region [2,21].

By using HA4R38A as the starting point for affinity maturation, we are decreasing the monobody's "preference" to bind in this pinching mode. It is then interesting to ask if we could mature HA4R38A back to a high-affinity clone, but one that binds in the typical way. In other words, can we force

protein binders to find alternative solutions (for example different epitopes) to the same target by taking a step back energetically and then proceeding with more affinity maturation. This would be interesting to follow up on using the mutagenesis strategy developed in the next two chapters.

There is almost overwhelming amount of choice available in the type of diversity to use in a protein library. There are many different ways that you could encode diversity into a protein binder library. Many parameters can be varied, such as the number of mutated residues within the scaffold of interest, the position of those residues relative to each other, the subset of the 20 amino acids that the mutated residues can possibly switch to, and even the probabilistic frequencies of mutation[*]. As a general rule, the designer does not know which library design strategy will work best a priori, and different antigens will be targeted optimally by different libraries.

The length of the loop libraries was one decision that was tricky. HA4's BC loop is 10 residues long. If we were only making slight differences to the amino acid sequence, e.g., only mutating 2 of the 10 residues at a time, it would make sense to keep it at 10. However the cassette mutagenesis strategy here completely replaces all the residues in the loop. The loops in this library would be completely different in amino acid character and would have different interactions with the surrounding protein structure. Wild-type ^{10}Fn3 domains tend to have 6, 7, 8, or 9, so we utilized those lengths as defaults.

Another parameter that must be chosen is the mutation profile. Oligo synthesis companies can insert degenerate nucleotides into your product, as was used in this work. These degenerate nucleotides represent a probabilistic mix of A,T,C, and G. The letter 'N' is used to denote a position in the oligo which could be any of the 4; during the oligo synthesis an equimolar mixture ("machine mixed") of all 4 phosphoramidite building blocks is allowed to react with the growing oligonucleotide polymer, giving a theoretical 25%[†] chance of any particular base being inserted. Whole codons can be constructed in this way, e.g. NNN is the codon where any of the 64 codons is equally probably.

In fact, even more advanced methods are available to more tightly define the library specificity. Oligonucleotide companies can use hand-mixed ratios of base reagents, instead of equimolar mixes, e.g.,

---

[*] E.g., 40% of my library members will have Ala at position 23 mutated to a Trp, 30% will have it mutated to Tyr, 25% to Phe, and 5% will have it remain Ala

[†] IDT's website claims that with the standard machine mixing option, skewing of the base ratios may occur due to "differences in delivery rate."

80% A, 10% T, 7% C, 3% G. As of this writing, an example 58-bp oligo with six hand-mixed nucleotides was five times as expensive as one using six machine-mixed nucleotides. When building multiple different libraries, this cost quickly becomes burdensome. An additional method[22], utilizing trinucleotide phosphoramidite reagents, can be employed to tightly control the amino acid frequency by building the oligo up in trinucleotide chunks instead of nucleotide by nucleotide. This method is also expensive.

In this work, we chose to stick with standard degenerate codons, as this significantly reduces the cost, and thus gives our mutagenesis technology much greater appeal and applicability. Given the observed bias in our library towards prolines, though, it is worth examining in the future where that source of codon bias arose from.

2.5 Materials and Methods

General molecular biology

Standard methods for molecular biology in *Saccharomyces cerevisiae* and *Escherichia coli* were used[23]. *S. cerevisiae* strains were grown at 30 °C in media containing 2% glucose unless otherwise noted. Cells were incubated in a New Brunswick Scientific Series 25 Incubator Shaker. Yeast cell density was determined by measuring optical density at 600 nm; OD = 1 was assumed to be a cell density of 1 x 10^7 cells per mL for strains EBY100, ysMDM100, and all their derivatives. Yeast electroporation was carried out using a previously reported protocol[24] on a Bio-Rad Gene Pulser Xcell. Yeast were stored by mixing yeast culture 1:1 with sterile 30% glycerol, dispensing in cryo-tubes, and storing at -80 °C. *E. coli* were stored long term by mixing liquid culture with 40% glycerol in a 1:1 proportion in cryotubes and then putting the tubes in a -80 °C freezer. *E. coli* transformation was carried out by electroporation using a Bio-Rad *E. coli* Pulser.

All aqueous solutions were made with distilled water prepared from a Milli-Q Water System. Restriction enzymes and Vent DNA polymerase were purchased from New England BioLabs, GoTaq DNA polymerase was purchased from Promega. dNTP mix was GE Healthcare illustra dNTPs purchased from Fisher Scientific. MJ Research PTC-200 Pellier Thermal Cyclers were employed for PCR. Plasmid DNA was purified from *E. coli* using QIAprep miniprep kits from Qiagen, for yeast minipreps, cells were

vortexed with acid-washed glass beads (Sigma) for five minutes before cell lysis. PCR products were purified with agarose gel electrophoresis and QIAquick spin columns. Agarose gels were run at 120-140 V using a Bio-Rad PowerPac 100. DNA was quantified using the UV absorbance at 260 nm on either a Molecular Devices SpectraMax Plus 384 instrument or a Tecan Infinite M200 with NanoQuant capabilities. Ethanol precipitations of DNA were done with Pellet Paint from Novagen. Plasmids and PCR products were eluted and stored in H_2O or EB buffer (10 mM Tris-HCl, pH 8.5) at -20 °C. Plasmid DNA was stored long-term by transformation into *E. coli* and then storing glycerol stocks of the cells at -80 °C.

All synthetic DNA oligonucleotides were purchased from Integrated DNA Technologies (Coralville, Iowa). PCR primers were typically designed to be 18-24 bp in length, with a T_M between 50 and 60 °C, and a GC content between 45 and 55%. T_M's were estimated on IDT's website http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/ which uses the nearest-neighbor two-state model of oligo thermodynamics with parameters taken from SantaLucia et al[25].

Routine DNA sequencing was carried out by mixing template DNA with 25 pmol of PCR primer in a PCR tube, and sending this sample mixture to GENEWIZ, Inc. (South Plainfield, NJ, www.genewiz.com). 96-well plate sequencing was done in the lab of Dr. Jingyue Ju at Columbia or by sending the plate to GENEWIZ.

Media and buffer recipes

- PBS (aka loading buffer): 20mM PO4, 150mM NaCl, pH 7.4
- PBSF buffer: Dissolve 8 g NaCl, 0.2 g KCl, 1.44 g Na_2HPO_4,0.24 g KH_2PO_4 and 1 g bovine serum albumin in 1 liter of deionized H_2O, adjust the pH to 7.4 and sterilize by filtration. This solution can be stored for up to 6 months at 4 °C (from Chao 2006).
- FPLC wash buffer: 20mM PO4, 150mM NaCl, 40 mM imidazole, pH 7.4
- FPLC elution buffer: 20mM PO4, 150mM NaCl, 500 mM imidazole, pH 7.4
- SG media (500 mL)
    - 250 mL 2x SD

- 25 mL of 20X amino acid mix
- 33.3 mL 30% galactose
- 192 mL of ddH_2O

DNA Ligation

Plasmids and PCR products were digested simultaneously by two different restriction enzymes (New England BioLabs) overnight. Buffers, incubation temperature, and inclusion of BSA were chosen to maximize enzyme activity. Double-cut DNA was purified through agarose gel electrophoresis, followed by band excision with a knife and gel extraction with Qiagen kit.

Ligation reactions were carried out in 20 µL volumes: 1 µL of T7 DNA ligase, 2 µL of 10x ligase buffer, 200-300 ng of digested plasmid, and X ng of insert, where X is chosen such that the molar ratio of insert to plasmid was between 10 and 30. The final volume was adjusted to 20 µL with ddH_2O. The reagents were mixed in a PCR tube, and kept at 16 °C for 10 hours on a thermocycler.

After this incubation, the reaction mix was pelleted with ethanol precipitation. Aliquots of electrocompetent TG1 *E. coli* cells were removed from -80 °C storage, thawed on ice for 10 minutes, and then 50 µL of cells were added to the pelleted ligation rxn and mixed by pipette. The cells were left on ice for an additional 10 minutes, transferred to an electroporation cuvette and then electroporated. Immediately after electroporation, the cells were resuspended in 1 mL of warm LB media and incubated at 37 °C for 1 hr. The cells were then plated on LB agar with appropriate antibiotic to select for the intended plasmid.

Making polyacrylamide gels for SDS-PAGE

10% gel (recipe below) was prepared and poured into gel cases. 500 µL isopropanol was added to smooth top of gel. The gel was allowed to harden for 45 min. The isopropanol was poured off and the the gel rinsed with ddH_2O. Stacking gel (recipe below) was prepared and added while inserting the lane comb in a chopping motion. The stacking gel was allowed to solidify for another 45 minutes.

10% Gel					**Stacking Gel**				
ddH_2O (mL)	2.47	4.94	9.88	14.82	ddH_2O (mL)	3.2	6.4	12.8	19.2
1.5 M Tris-Cl/SDS pH 8.8 (mL)	1.25	2.5	5	7.5	5 M Tris-Cl/SDS pH 6.8 (mL)	1.25	2.5	5	7.5
Acrylamide/Bis 40% (mL)	1.25	2.5	5	7.5	Acrylamide/Bis 40% (mL)	0.5	1	2	3
10% APS (µL)	25	50	100	150	10% APS (µL)	25	50	100	150
TEMED (µL)	3.75	7.5	15	22.5	TEMED (µL)	5	10	20	30
Total Volume (mL)	5	10	20	30	Total Volume (mL)	5	10	20	30

E. coli colony PCR

Colonies from an agar plate were picked up with pipette tips and suspended in 20 µL of H_2O in a PCR well (optionally the pipette tips can then be used to patch a replica agar plate for further storage of the colony). A PCR mix was made with 1 µL of 10 µM forward primer, 1 µL of 10 µM reverse primer, 0.5 µL of dNTP mix, 4 µL of 5x GoTaq Buffer, 3 µL of H_2O, and 0.5 µL GoTaq polymerase (purchased from Promega). 10 µL of this PCR mix was added to 10 µL of the cell suspension and program EPP5 was run on the thermocycler. Agarose gel electrophoresis was used for quality inspection of PCR products and for gel purification of bands. Bands purified by Qiagen prep kit were sent to GENEWIZ for sequencing.

Expression of GST-SH2 fusion protein

Plasmid pGST-SH2 was obtained from the Koide Lab at the University of Chicago. Competent BL21 DE3 *E. coli* cells were thawed from the freezer and added to the 2 µL of plasmid. Transformation was carried out by electroporation. Cells were grown on LB + kanamycin plates.

The following protein expression protocol follows the auto-induction method of Studier, see his paper[6] for all media recipes and protocol details. A colony from the LB+kanamycin plates was inoculated

into 5 mL P-0.5G media in a 25 mL flask, which was put shaking in 37 °C shaker overnight. The following afternoon, the overnight culture was inoculated 1:1000 into ZYP-5052 auto-induction media (250 mL – 2L depending on amount of protein desired) and grown overnight at 30 °C. The following morning, 20 µL of each flask was taken for SDS-PAGE. The rest of the cultures were transferred to sterile centrifuge bottles (250 mL / bottle) and spun 10,000xg, 15 min, 4 °C on Beckman Coulter Avanti J-E Centrifuge. The supernatant was poured off and the cells were resuspended in 40 mL of 20mM Phosphate buffer, 150mM NaCl with 5mM beta mercaptoethanol (pH 7.4) per 500 mL of original culture volume. Resuspension was carried out by pipetting up and down until all the cells were thoroughly separated. The resuspended cells were frozen at -20 ºC; this step helps break open the cell membranes. The cell pellet was thawed and PMSF was added to a final concentration of 1 µM. Then hen egg lysozyme was added to a final concentration of 0.5-1 mg/mL and the cell suspension was mixed thoroughly by swirling. The mixture was incubated at 37 °C 30 min.

150 µL of cell lysate was pipette up and down to get a sense of the viscosity. The cells were transferred to a glass beaker and sonicated (QSONICA Misonix S-4000) on ice at a power of 50 for 4-6 cycles of 30 sec on, 30 sec off. To verify that sonication had successfully lysed the cell membranes, 150 µL of cell suspension was pipetted up and down, and compared to the pre-sonication viscosity. The cell lysate should be watery (low-viscosity) at this point. 20 µL of this lysate was taken for SDS-PAGE. The rest of the lysate was centrifuged at 30,000xg for 15 min at 4 ºC. The supernatant was collected for column purification. 4 µL of 6X SDS was added to each 20 µL sample, and boiled on a hot plate for 10 min (95 °C). Samples were run on 10% polyacrylamide gels at 120 V for 1 – 1.5 hrs, with Precision Plus Protein™ Kaleidoscope™ standards from Bio-Rad.

Purification of GST-SH2 fusion protein

All affinity column purifications were performed on a GE Amersham Pharmacia Biotech AKTAFPLC (P-920 and UPC-900) using GE Histrap HP columns (1mL or 5 mL).

10 column volumes of H_2O were run to clear the storage liquid out of the system and the injection loop. Then 10 column volumes of PBS (20mM PO4, 150mM NaCl, pH 7.4) were run to equilibrate the column

and fill the injection loop. The GST-SH2 supernatant was filtered through 0.22 µm Pall Acrodisc filters. Using a 12 mL syringe and an 20 g needle, 10 mL of filtered protein solution was slowly pulled up, so as to avoid formation of bubbles, and then loaded into the injection loop. The protein solution within the loop was then pushed onto the column with 10 mL of loading buffer, collecting fractions as "flow-through." Protein was eluted by running a linear gradient over 20 column volumes from 10% elution buffer to 100%, where elution buffer is 20mM PO4, 150mM NaCl, 500 mM imidazole, pH 7.4.

Biotinylation of GST-SH2

To remove the imidazole and reduce the cysteines with DTT, buffer exchange was carried out with Amicon spin filters with a molecular weight cut-off of 10,000 daltons.

Clean fractions (as determined by SDS-PAGE) from column purification were combined into one protein solution. DTT was added to the protein solution to a final concentration of 5 mM, this was mixed on a rotator for 10 min. Then the protein solution was added to an Amicon filter pre-equilibrated with PBS. The solution was centrifuged at 3 krpm until concentrate was < 1 mL (roughly 90 minutes). PBS with 5 mM EDTA was added up to 15 mL. Two more repetitions of centrifugation and dilution with PBS/EDTA were carried out. Final protein concentration was then quantitated via A_{280} on a Molecular Devices SpectraMax Plus 384 instrument or a Tecan Infinite M200[26]. Using the primary amino acid sequence and the ProtParam bioinformatics tool at ExPASy (www.expasy.org), the extinction coefficient of the translated GST-SH2 protein product was calculated to be 64,750 $M^{-1}cm^{-1}$, with molecular weight 43,100 g/mol.

To make HPDP stock, 3.1 mg of EZ-Link HPDP-Biotin (Thermo Scientific) was added to 1 mL DMSO in a glass vial for a final concentration of 5.6 mM. 1.25 mL of SH2-GST solution from spin filter and 100 µL of HPDP stock were combined in a microfuge tube, and put rotating at RT for 2 hrs. Using a fresh spin filter or dialysis, the unreacted HPDP reagent was buffer exchanged away with 3x 15 mL of PBS.

Quantitation of biotin was carried out using a Pierce Biotin Quantitation Kit (Thermo Scientific). 800 µL of PBS was added to a cuvette. This was used to blank a spectrophotometer at wavelength 500 nm. 100 µL of a mixture of 4'-hydroxyazobenzene-2-carboxylic acid (HABA) and avidin were added to the

cuvette, and an absorbance measurement at 500 nm was taken as a starting point for the assay. To start the assay, 100 µL of biotinylated GST-SH2 was added to the cuvette. Every ten seconds, an absorbance measurement was taken. When the absorbance reading was the same for at least 30 sec, that reading was taken as the endpoint and used to calculate the molar ratio of biotin

For storage, the biotinylated GST-SH2 was aliquoted into individual microfuge tubes and flash frozen in liquid nitrogen, and then stored at -80 °C.

Galactose Induction of Yeast Display

The library of cells was grown in selective SC media to OD_{600} between 1 and 3. The desired number of cells for selection was transferred to a 50 mL conical tube and spun down 15 min at 2800xg. The supernatant was poured off by a sterile flame, and the cells were resuspended in selective SG media to a final OD between 0.5 and 1. The conical tubes were then put on a rotator at 23-25 °C for 20-36 hrs.

Magnetic Bead Capacity Experiment

I subcloned the HA4 and HA4R38A genes into the yeast display vector within EBY100 cells by gap-repair. For the purpose of having a strong positive control, I obtained from the Wittrup lab a yeast display plasmid (pCT-L7.5.1) for a fibronectin binder (L7.5.1) with K_d = 3 pM and transformed this into EBY100, as well[8]. The target of binding of L7.5.1 is lysozyme, and Lysozyme Biotin-Caproyl was purchased from Sigma-Aldrich. For a negative control, I transformed pGalR38A, which has a non-sense stop codon in the middle of the BC loop, into EBY100, to make ysEBYR38A*.

Dynabeads M-280 streptavidin-coated beads and magnetic rack (MagnaRack CS15000) were purchased from Invitrogen. Settled beads stored at 4 °C were resuspended by vortexing 2-5 sec, and 10 µL of bead suspension ($5x10^6$ beads) were transferred to individual 1.5 mL microfuge tubes and washed 3x with 1 mL PBSF using the magnetic rack to separate beads from supernatant. The beads were coated with 100 µL of 36.7 pmol biotinylated target, either GST-SH2 for ysEBYHA4 or lysozyme for ysEBYL751, and then left rotating overnight at 4 °C.

SC dropout cultures of strains ysEBYHA4, ysEBYL751, and ysEBYR38A* were grown to saturation (OD_{600} = 3 – 4). On the day before selection, 25 mL of culture were transferred to 50 mL conical tubes and centrifuged 2800xg for 10 min. The supernatant was poured off, and the cell pellets were resuspended in 25 mL of SG dropout media. These induction cultures were put on the rotator for 24 hrs at 23 °C.

On the day of selection, $5x10^8$ cells were taken from each galactose induction culture and spun down 2800xg for 5 min, and the supernatant poured off. In a cold room maintained at 4 °C, the beads were pulled down magnetically and the supernatant pipetted off. The beads were washed 2x with 1 mL cold PBSF in the same manner. The galactose-induced cells were resuspended in 1 mL cold PBSF and added to the target-coated beads in the microfuge tubes. The cell-bead mixtures were put on the rotator at 4 °C for 2.5 hrs.

After this incubation, the tubes were put on the magnet for 5 min. The sup was pulled off and dilution plated on SC(UT-) dropout plates to determine # of cells not bound to beads. The beads were washed 4x with 1 mL cold PBSF. The beads were then resuspended in 1 mL PBSF and dilution plated on SC(UT-) to determine # of cells bound to beads.

Magnetic Bead Mock Library Enrichment

Galactose inductions were carried out as above; SG(UT-) media was used for strains ysEBYHA4, ysEBYL751, ysEBYR38A*, and SG(U-) media for strain EBY100 (containing no display plasmid). 2.5 x 10^6 magnetic beads were labeled overnight at 4 °C as above.

On the day of the selection, $5x10^5$ cells of each monobody display strain were mixed with $5x10^8$ EBY100 (background) cells immediately prior to adding the cells to the washed magnetic beads. Dilution plates were made at this point to determine the pre-selection library ratio, with SC(U-) plates for growth of all cells and SC(UT-) plates for growth of cells containing display plasmid. The library-bead mixtures were than rotated at 4 °C for 2 hrs 40 min After rotation, the beads were washed 4x with 1 mL cold PBSF. The beads were resuspended in 1 mL PBSF and dilution plates were made to determine the post-selection library ratio.

Flow Cytometry of ysEBYHA4

7.7 µL bead slurry (5 x 10^6 magnetic beads) were added to each tube, washed 2x with PBSF and incubated with 100 µL of 367 nM GST-SH2 overnight. 10^6 ysEBYHA4 and EBY100 cells from galactose induction were centrifuged at 14 krpm for 1 min. The cells were washed once with 1 mL PBSF, and then resuspended in 50 µL of 23.4 nM SH2-GST-Biotin and then put in the cold room to rotate for 3.5 hrs. The rest of the steps were done on ice. The cells were pelleted by spinning 30 sec @ 14,000 rpm. Supernatant was pipetted off and the cells were washed with ice-cold 1 mL PBSF. Cells were then resuspended in 50 µL of PBSF, to which 0.5 µL of FITC-anti-V5 and 0.5 µL of PE-streptavidin were added. These were rotated at 4 °C for 10-20 min. The cells were then spun 30 sec @ 14,000xg and washed with ice-cold 1 mL PBSF. The supernatant was removed and the cell pellets were left on ice during physical transit to the flow cytometry core at Columbia University Medical Center. Immediately prior to cytometry, the cell pellets were resuspended in 500-800 µL of cold PBSF. Cytometry measurements were made on a BD FACSCanto analyzer.

Construction of yeast display plasmid for homologous recombination experiments

The backbone for yeast display is the pYD1 plasmid, and the yeast strain used is EBY100. The earliest libraries were created by gap-repair. In gap-repair editing of plasmids, the backbone plasmid is first linearized with restriction enzymes, ideally with the cut site being in the center of the DNA region you wish to edit. An insert is produced by PCR or polymerase extension of synthetic oligonucletides such that both ends of the insert have homology (generally > 30 bp) to the plasmid on both sides of the cut site made above. The insert and the cut plasmid are then co-transformed into yeast. Once inside the cell, the yeast homologous recombination machinery replaces the cut DNA in the plasmid with the insert, and in doing so, re-circularizes the plasmid.

I knew early on that I wanted to utilize homologous recombination to replace one of the ^{10}Fn3 loops with a loop library. In order to assess the success of the technology as I developed it, I wanted to build a plasmid that would be non-functional until homologous recombination with my libraries took place.

Furthermore, I wanted the process of homologous recombination to remove a unique cut-site; the presence of cutting discernable on an agarose gel could then serve as a diagnostic of whether recombination took place or note. To accomplish both of these goals, I created pGalEng (for *eng*ineered BC loop), an ^{10}Fn3 displaying vector where the wt BC loop was replaced with one containing a non-sense stop codon and a ClaI cut site; I call the resulting sequence the "template" BC loop.

Since I had used pGalAgaFNC as the backbone, I now had a hybrid ^{10}Fn3 gene with the N-terminal part from HA4R38A, the template BC loop, and the wild-type fibronectin C-terminus. To finish, I had to subclone the C-terminal part of HA4R38A into pGalEng. This gave me pGalR38A, which was the starting point for my library mutations.

pGalEng was constructed by first digesting pGalAgaFNC (a gift from the Koide Lab) with NheI and EcoRI. A 5' insert fragment was made by PCR of the pHFT2-R38A plasmid with primers MDM0015 and MDM0013, and a 3' insert fragment was made by PCR of pHFT2-R38A with primers MDM0025 and MDM0023. The linearized pGalAgaFNC and the two insert fragments were all co-transformed into yeast, with gap-repair resulting in the desired product.

To construct pGalR38A, primers MDM0055 and MDM0053 were used to PCR the C-terminal end of the HA4R38A gene in pHFT2-R38A. pGalEng was digested at a single site with SacI, and the digested enzyme was co-transformed with either the HA4 C-terminal insert or the R38A insert into yeast, where gap repair by homologous recombination yielded the final product.

Sequencing of ^{10}Fn3 in pGalEng and in all later yeast display vectors was carried out with oligos MDM0035 and MDM0032.

The χ^2 test on 8x NNB library was performed using 18 bins (the six categories with predicted freq. 2% were paired into three bins to meet the criteria for using the test) and thus 17 degrees of freedom.

Construction of mutagenic BC cassettes

The construction of the BC loop libraries is illustrated in Figure 2-X. These libraries were constructed by overlap extension PCR.

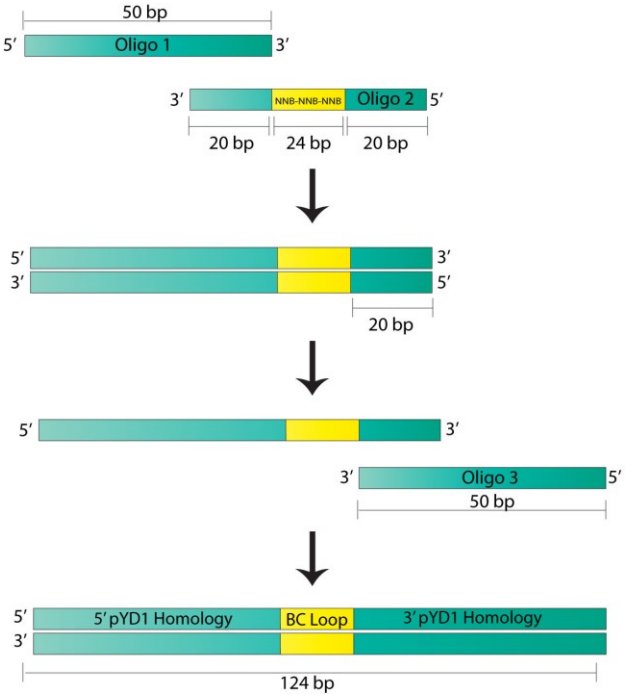


Figure 2-8. Construction of mutagenic BC cassettes by oligonucleotide extension PCR. Each rectangle represents a single-stranded stretch of DNA. The stretch of randomized codons is shown in yellow.

The reaction was carried out in a single PCR reaction, for which the recipe was:

1 µL of 100 µM oligo 1 (MDM0041)

1 µL of 100 µM oligo 3 (MDM0043)

1 µL of 10 µM oligo 2 (either one of or a mixture of the following oligos)

- MDM042-2 for 6 NNB codons
- MDM042-1 for 7 NNB codons
- MDM0042 for 8 NNB codons
- MDM042+1 for 9 NNB codons
- MDM045-2 for 6 NNY codons
- MDM045-1 for 7 NNY codons
- MDM0045 for 8 NNY codons
- MDM045+2 for 9 NNY codons

1 µL of dNTP mix

5 µL of 10x pfu buffer

40.5 µL of ddH_2O

0.5 µL of PfuTurbo DNA polymerase (from Stratagene)

The thermocycler program was:

The Koide Lab at the University of Chicago sent us pHFT2 plasmids containing the ^{10}Fn3 mutants HA4 (K_d = 10 nM), HA4R38A (K_d ≈ 350 nM), and the Abl1 SH2 domain. The pHFT2 vector is based on pET, contains a kanamycin resistance gene, an N-terminal 10-His tag cleavable with TEV protease.

FACS sorting of libraries

FACS was carried out as above for ysEBYHA4. After galactose induction of libraries, 7.5 x 10^7 cells from each library were spun down, washed with 1 mL PBSF, and then resuspended in 1 mL PBSF. Sorting was carried out on a BD Influx™ cell sorter.

Yeast Transformation

The night before electroporation, yeast cells were inoculated from a fresh patch onto a culture plate in 5 mL of selective complete (SC) medium. Shook cells at 30° C O/N

Put 100 mL of YPD in separate flask and warm in shaker O/N, as well

The next morning, the OD600 of a 1/10 dilution of O/N culture was determined. The O/N culture was inoculated into 100 mL of YPD to an OD=0.1 The cells were grown in a 30°C shaker until OD600≈0.6-0.8 (usually after 4-5 hrs.).

Transformation DNA was distributed into 6 different microcentrifuge tubes and pelleted.

- Tube 1: no DNA
- Tube 2: cut vector only, 0.5 µg
- Tubes 3-6: 3 µg cut vector + 10 µg library PCR fragments (Wittrup)

1ml of filter-sterilized 1,4-dithiothreitol (DTT) solution (1M Tris, pH 8.0, 2.5M DTT) was added to the 100ml of YPD. The cells were allowed to grow for another 20 min at 30°C. The Tris was prepared ahead of time, and then the DTT was added to the Tris right before adding it to the YPD.

From this point, the cells were kept on ice. The cells were centrifuged at 5min, 2000rpm, 4C, in 50ml falcon tubes. The supernatant was discarded and the cells were washed with 25ml of cold E buffer

(10mM Tris, pH 7.5, 270mM sucrose, 1mM $MgCl_2$) by pipetting up and down. The cells were spun down again, with the supernatant discarded, and the cells were resuspended in 1ml of E buffer and transfered to a 1.5ml microcentrifuge tube and spun down again (10sec total, 12000rpm). The supernatant was discarded and the cells were resuspended with 50 µl E-buffer. Distributed 50 µL cells into each of the 6 microcentrifuge tubes containing the pelleted DNA.

The cells and pelleted DNA were mixed by pipetting until the pellet was no longer visible, and then the mixture was allowed to sit for 10 min on ice. The electroporator was set to 25 µF, 540 V, 0 ohms. The cells were transferred to a 0.2cm eletroporation cuvette and pulsed. Immediately after electroporation, 1ml warm YPD was added to the cells, which were then shook at 30ºC for 1hr.

After incubation, the cells were transfered to microcentrifuge tubes and spun at 10k rpm for 10 sec at R.T. The cell pellets were resuspended in 200 µL of sterile water. Tubes 1 and 2 (negative controls) were plated on selective media to determine degree of background growth. Tubes 3-6 were pooled and and transfer to 200 mL of selective media. 8 µL of this mixture was removed for dilution plating to determine library size. The liquid cultures and plates were incubated at 30 °C for 2 days

After incubation, colonies on plates were counted to determine library size and background growth rate. The 200 mL culture was centrifuged at 3000 g for 5 min and supernatant was discarded. The cells were resuspended in 200 mL sterile water. Centrifuge was carried out again at 3000 g for 5 min. The cell pellets were resuspended in 7.5 mL sterile water, and 5 µL were taken to make a 200x dilution for OD600 determination. OD600 = 1 $\Rightarrow$ 10^7 cells/mL. $1x10^9$ – $1x10^{10}$ cells/mL (A600 = 100-1000) are expected. Glycerol stocks were then made using at least 20X library size per tube.

Yeast Colony PCR

The following is adapted from a protocol from the Hahn Lab: http://labs.fhcrc.org/hahn/Methods/mol_bio_meth/pcr_yeast_colony.html

Colonies from a media plate were picked up with a pipette tip or inoculating loop and suspended in 30 µL 0.2% SDS in individual PCR tubes. The SDS suspensions were vortexed on high for 30 sec, and then the PCR tubes were heated to 95 °C for 5 min on a thermocycler. The tubes were then spun on a

centrifuge @ 2800xg for 5 min. 0.5 µL of the supernatant was added to a 50 µL total PCR reaction mix, which also included 0.5 µL dNTP mix, 0.5 µL MDM0035, 0.5 µL MDM0032, 10 µL of 5x GoTaq Buffer, 37.5 µL of H2O, and 0.5 µL of GoTaq DNA polymerase.

This reaction was cycled 25 times, with 54 °C annealing temperature. After thermocycling, the PCR products were purified by agarose gel electrophoresis followed by gel extraction with a QIAquick Gel Extraction kit (Qiagen). Alternatively, if high throughput number of colonies were being sequenced, the PCR reaction would be purified enzymatically by adding 1 µL of ExoSAP-IT to 9 µL of PCR reaction followed by 15 min incubation at 37 °C, then 15 min incubation at 80 °C. With either method, 10 µL of purified PCR product were mixed with 5 µL of 5 µM PCR primer and sent to GENEWIZ for sequencing.

2.6 Strains, Plasmids, and Oligonucleotides

Table 2-7. Strains used in this study

Name	Details	Source/Reference
EBY100	MAT**a** *ura3-52 trp1 leu2Δ1 his3Δ200 pep4:HIS3 prb1Δ1.6R can1 GAL* yeast display Aga system	Koide Lab/[1]
ysEBYHA4	EBY100 containing pYD-HA4	
ysEBYL751	EBY100 containing pCT-L7.5.1	
ysEBYR38A*	EBY100 containing pGalR38A	

Table 2-8. Plasmids used in this study

Name	Details	Source/Reference
pYD1	pUC-ori P_{GAL1} TRP1 CEN6/ARSH4 amp^R	Koide Lab
pGalAgaFNC	pUC-ori P_{GAL1} N-term ^{10}Fn3 TRP1 CEN6/ARSH4 amp^R	Koide Lab
pGST-SH2	GST-SH2 fusion protein under T7 *lac* promoter kan^R	Koide Lab/1
pGalEng	pGalAgaFNC with template BC loop (ClaI and stop)	
pGalR38A	pYD1 containing full HA438A with template BC	pGalEng
pYD-HA4	pYD1 containing HA4 monobody	
pCT-L7.5.1	Yeast diplay plasmid containing Wittrup pM binder	Wittrup Lab/[8]

Table 2-9. Oligonucleotides used in this study

Name	Sequence
MDM0015	GGTGGTGGTGGTTCTGCTAGCGGTGGCGGTGTTTCTTCTGTTCCGACCAA
MDM0013	ATGGCGTAATACTAAACAGAAGAAGAAGACATCGATGCATCCCAGCTGAT
MDM0025	ATCAGCTGGGATGCATCGATGTCTTCTTCTTCTGTTTAGTATTACGCCAT
MDM0023	GTACAGTGAATTCCTGAACCGG
MDM0035	CCGGATCGGACTACTAGC
MDM0032	ACTGTTGTTATCAGATCAGCGG
MDM0041	CCAAACTGGAAGTTGTTGCTGCGACCCCGACTAGCCTGCTGATCAGCTGG
MDM0042	CCGTACGTGATGGCGTAATAVNNVNNVNNVNNVNNVNNVNNVNNCCAGCTGATCAGCAGGCTAG
MDM042-2	CCGTACGTGATGGCGTAATAVNNVNNVNNVNNVNNVNNCCAGCTGATCAGCAGGCTAG
MDM042-1	CCGTACGTGATGGCGTAATAVNNVNNVNNVNNVNNVNNVNNCCAGCTGATCAGCAGGCTAG
MDM042+1	CCGTACGTGATGGCGTAATAVNNVNNVNNVNNVNNVNNVNNVNNCCAGCTGATCAGCAGGCTAG
MDM0043	TCCTGAACCGGGGAGTTACCACCGGTTTCACCGTACGTGATGGCGTAATA

MDM0045	CCGTACGTGATGGCGTAATARNNRNNRNNRNNRNNRNNRNNRNNCCAGCTGATCAGC AGGCTAG
MDM045-2	CCGTACGTGATGGCGTAATARNNRNNRNNRNNRNNRNNCCAGCTGATCAGCAGGCTA G
MDM045-1	CCGTACGTGATGGCGTAATARNNRNNRNNRNNRNNRNNRNNCCAGCTGATCAGCAGG CTAG
MDM045+1	CCGTACGTGATGGCGTAATARNNRNNRNNRNNRNNRNNRNNRNNRNNCCAGCTGATC AGCAGGCTAG
MDM0055	CATCACGTACGGTGAAACC
MDM0053	GTACGGTAGTTAATCGAGATTGG

2.7 References

1. Boder, E. T. & Wittrup, K. D. Yeast surface display for screening combinatorial polypeptide libraries. *Nature biotechnology* **15**, 553–557 (1997).

2. Wojcik, J. *et al.* A potent and highly specific FN3 monobody inhibitor of the Abl SH2 domain. *Nat Struct Mol Biol* **17**, 519–527 (2010).

3. Ackerman, M. *et al.* Highly avid magnetic bead capture: An efficient selection method for de novo protein engineering utilizing yeast surface display. *Biotechnology Progress* **25**, 774–783 (2009).

4. Boder, E. T. & Wittrup, K. D. Optimal Screening of Surface-Displayed Polypeptide Libraries. *Biotechnology Progress* **14**, 55–62 (1998).

5. Gera, N., Hussain, M. & Rao, B. M. Protein selection using yeast surface display. *Methods* (2012).at <http://www.sciencedirect.com/science/article/pii/S1046202312000709>

6. Studier, F. W. Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification* **41**, 207–234 (2005).

7. Michael Green, N. [5] Avidin and streptavidin. *Methods in Enzymology* **Volume 184**, 51–67 (1990).

8. Hackel, B. J., Kapila, A. & Dane Wittrup, K. Picomolar Affinity Fibronectin Domains Engineered Utilizing Loop Length Diversity, Recursive Mutagenesis, and Loop Shuffling. *Journal of Molecular Biology* **381**, 1238–1252 (2008).

9. Feldhaus, M. J. *et al.* Flow-cytometric isolation of human antibodies from a nonimmune Saccharomyces cerevisiae surface display library. *Nature biotechnology* **21**, 163–170 (2003).

10. Swers, J. S., Kellogg, B. A. & Wittrup, K. D. Shuffled antibody libraries created by in vivo homologous recombination and yeast surface display. *Nucleic Acids Research* **32**, e36– (2004).

11. Koide, A., Gilbreth, R. N., Esaki, K., Tereshko, V. & Koide, S. High-affinity single-domain binding proteins with a binary-code interface. *Proceedings of the National Academy of Sciences, USA* **104**, 6632–6637 (2007).

12. Fellouse, F. A. *et al.* High-throughput generation of synthetic antibodies from highly functional minimalist phage-displayed libraries. *Journal of Molecular Biology* **373**, 924–940 (2007).

13. Birtalan, S. *et al.* The Intrinsic Contributions of Tyrosine, Serine, Glycine and Arginine to the Affinity and Specificity of Antibodies. *Journal of Molecular Biology* **377**, 1518–1528 (2008).

14. Lipovsek, D. *et al.* Evolution of an Interloop Disulfide Bond in High-Affinity Antibody Mimics Based on Fibronectin Type III Domain and Selected by Yeast Surface Display: Molecular Convergence with Single-Domain Camelid and Shark Antibodies. *Journal of Molecular Biology* **368**, 1024–1041 (2007).

15. Reetz, M. T., Kahakeaw, D. & Lohmer, R. Addressing the Numbers Problem in Directed Evolution. *ChemBioChem* **9**, 1797–1804 (2008).

16. Wong, T. S., Roccatano, D. & Schwaneberg, U. Steering directed protein evolution: strategies to manage combinatorial complexity of mutant libraries. *Environmental Microbiology* **9**, 2645–2659 (2007).

17. Gilbreth, R. N., Esaki, K., Koide, A., Sidhu, S. S. & Koide, S. A Dominant Conformational Role for Amino Acid Diversity in Minimalist Protein-Protein Interfaces. *Journal of Molecular Biology* **381**, 407–418 (2008).

18. Firth, A. E. & Patrick, W. M. GLUE-IT and PEDEL-AA: new programmes for analyzing protein diversity in randomized libraries. *Nucleic Acids Research* **36**, W281–285 (2008).

19. Patrick, W. M. & Firth, A. E. Strategies and computational tools for improving randomized protein libraries. *Biomolecular Engineering* **22**, 105–112 (2005).

20. Lee, C. V *et al.* High-affinity Human Antibodies from Phage-displayed Synthetic Fab Libraries with a Single Framework Scaffold. *Journal of Molecular Biology* **340**, 1073–1093 (2004).

21. Koide, A., Wojcik, J., Gilbreth, R. N., Hoey, R. J. & Koide, S. Teaching an Old Scaffold New Tricks: Monobodies Constructed Using Alternative Surfaces of the FN3 Scaffold. *Journal of Molecular Biology* (2012).at <http://www.scopus.com/inward/record.url?eid=2-s2.0-84455161592&partnerID=40&md5=aca2d7155e2223743b803d424349e9a8>

22. Virnekas, B. *et al.* Trinucleotide phosphoramidites: ideal reagents for the synthesis of mixed oligonucleotides for random mutagenesis. *Nucleic Acids Research* **22**, 5600–5607 (1994).

23. *Current Protocols in Molecular Biology*. (John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2001).doi:10.1002/0471142727

24. Pirakitikulr, N., Ostrov, N., Peralta-Yahya, P. & Cornish, V. W. PCRless library mutagenesis via oligonucleotide recombination in yeast. *Protein Science* n/a–n/a (2010).doi:10.1002/pro.513

25. SantaLucia John, Allawi, H. T. & Seneviratne, P. A. Improved Nearest-Neighbor Parameters for Predicting DNA Duplex Stability. *Biochemistry* **35**, 3555–3562 (1996).

26. Noble, J. E. & Bailey, M. J. A. Quantitation of Protein. *Methods in Enzymology, Vol 463: Guide to Protein Purification* **Volume 463**, 73–95 (2009).

Chapter 3

A recombination system for *in vivo* mutagenesis

3.1 Abstract

In order to create libraries through *in vivo* mutagenesis, we need to have a system that can transfer genetic material from a cassette plasmid to a target plasmid through homologous recombination. Here, we describe the construction of the three plasmid components and the editing of the yeast genome in order to effect efficient cassette-based mutagenesis. Assays for homologous recombination demonstrate genetic conversion of the DNA encoding the BC loop in the monobody gene. A timecourse induction with doxycycline showed that significant levels of homologous recombination are detectable after 3-4 days of induction under a strong $tetO_7$ promoter. Finally, induction and selection of mock libraries demonstrated that recombinants could be enriched by magnetic beads, suggesting that their reconstituted HA4 monobody was functional.

3.2 Introduction

The main objective of the work described in this chapter was to build and assemble the genetic components required for *in vivo* mutagenesis, and then to confirm that DNA from a cassette plasmid was being successfully recombined into the yeast display plasmid. Because plasmids can be passed from one cell to another by mating, such a system would allow us to pass mutational information horizontally within a population.

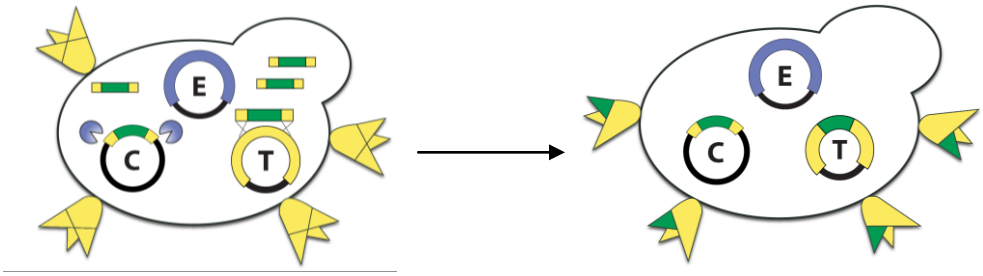


Figure 3-1. A three plasmid system for *in vivo* mutagenesis. The wild-type monobody gene (yellow) is contained within the yeast display plasmid, **T**. An endonuclease is encoded on a second plasmid, **E**. A DNA cassette is contained on a third plasmid, **C**. The cassette contains a mutant sequence (green) for replacing a wild-type sequence in the monobody, flanked on both ends by homologous DNA which allows the cassette to be targeted to the monobody gene on plasmid **T**. Induction of the endonuclease enzyme (blue pacman) followed by double-strand break formation and homologous recombination results in a mutated display plasmid **T** where the wild-type DNA has been replaced by the cassette. The resulting protein monobodies are show expressed on the cell surface (yellow and green floret).

The key to our mutagenesis system is the use of homologous recombination to replace DNA in a target gene with a mutant sequence obtained from a separate plasmid. Our strategy for accomplishing this mutagenesis requires a 3-plasmid genetic system (Figure 3-1). One plasmid contains the monobody gene and is responsible for expressing the monobody protein on the surface of the yeast. A second plasmid contains the mutagenic DNA cassette. A third plasmid expresses an endonuclease which generates double-stranded breaks (DSB's) in the DNA. Ideally, we wish to have a system such that any yeast cell containing both our template plasmid and our cassette plasmid will rapidly and accurately replace the sequence of interest in the template plasmid with the mutant sequence from the cassette plasmid.

To drive homologous recombination (and thus mutagenesis in our system), we needed to introduce double-stranded breaks (DSB's) into the DNA *in vivo*[5]. *Saccharomyces cerevisiae* is very efficient at repairing double-strand breaks (DSB's) using homologous DNA[6,7]. They have a complex assortment of cellular machinery for detecting and fixing DSB's[8,9]. These proteins are responsible for processing the cut DNA ends and searching through all of the DNA in the nucleus for sequence homology. Both the yeast chromosomal genome and any plasmids that happen to be present in the nucleus can be potential donors of homologous DNA. Once homologous DNA is found, the

recombination machinery aids a DNA polymerase in repairing the DSB through new DNA synthesis, using the homologous sequence from the donor as a template.

We chose to use the I-SceI enzyme to generate the DSB's in vivo. I-SceI is an endonuclease that is expressed naturally in the yeast mitochondria. It generates DSB's at highly specific 18 bp recognition sequences[10]. Plessis et al. demonstrated that I-SceI expressed from a plasmid could be active in the yeast nucleus and catalyze double-strand breaks at its specific cut-site[4].

The basic idea (outlined in Chapter 1, Figure 1-3) is as follows. We use the highly specific I-SceI endonuclease to make a double-stranded break (DSB) immediately adjacent to a mutagenic cassette. The yeast homologous recombination machinery is recruited to this cut. The homologous region next to the cut is then used to target the cassette to the monobody gene in the yeast display plasmid. Homologous recombination results in the conversion of the existing BC loop sequence in the display plasmid to the mutant sequence encoded by the cassette plasmid. Induction of yeast display with galactose puts this mutated protein variant on the surface of the yeast, and phenotypic selection by magnetic beads or FACS allows us to isolate the cells with the most functional mutants. These high-functioning yeast can then be mated with a library of yeast containing a different FG loop variants, resulting in a mixing of mutagenic cassette plasmids which target the two different loops.

3.3 Results and Discussion

Construction of diagnostic yeast display plasmid

We built a yeast display plasmid that contained the HA4 gene with a mutated BC loop. This plasmid (named pHRS-YD1B) contains a BC loop that was intentionally altered from the HA4 BC in 3 ways: 1) its sequence is DASMTWLSV* where * is an Amber TAG stop codon, 2) it contains a ClaI cut site, and 3) it lacks a flanking BsiWI cut site (Figure 3-2). The purpose of pHRS-YD1B and its "placeholder" BC loop is to serve as a diagnostic for whether homologous recombination with the mutagenic cassette has taken place or not. The fixing cassette (in the cassette plasmid) was designed to replace this placeholder BC loop with the wild-type HA4 BC sequence (DAPMSSSSVY), which 1) fixes the stop codon, 2) removes the ClaI site, and 3) adds a BsiWI site. Because a V5 epitope tag is located

C-terminal to the monobody, labeling with fluorescent anti-V5 antibody followed by flow cytometry enables us to differentiate between cells with unrecombined plasmid (stop codon) and cells with recombined plasmids (expressed V5). This was all intended so that I would be able to detect homologous recombination and thus gauge the effectiveness of the mutagenesis system.

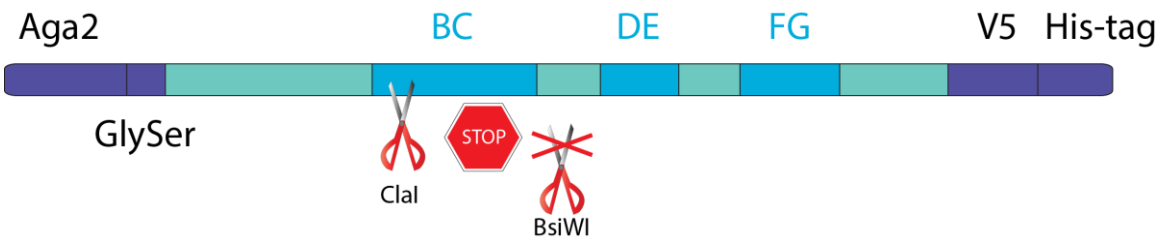


Figure 3-2. BC placeholder in pHRS-YD1B has diagnostic features. Yeast display plasmid pHRS-YD1B contains a modified HA4 gene (scaffold framework regions in teal, loop regions in blue). A stop codon has been added to the BC so that unrecombined plasmids produce a truncated protein on the yeast surface without the downstream V5 epitope. A ClaI cut site has been added (scissors) and a BsiWI cut site has been removed with a silent mutation (crossed out scissors). Thus unrecombined plasmid cuts with ClaI and doesn't cut with BsiWI.

Plasmids for inducible expression of I-SceI endonuclease and DSB's

The gene coding for the I-SceI endonuclease was cloned into CEN plasmids under the control of the reverse tetracycline inducible expression system[11,12]. Under this system, addition of doxycycline to the cell medium activates expression of I-SceI. We chose such an inducible expression system because we only want DSB's to be formed during a well-defined induction time. Constitutive expression of I-SceI would be expected to lead to genetic instability of the cassette plasmids. Additionally, the constant presence of DNA DSB's would be expected to slow growth of cells via cell cycle arrest[13], which could result in significant growth biases within a population of cells.

The strength of the activation depends upon the number of regulatory tetO boxes present upstream of the I-SceI gene. We cloned the I-SceI gene into two expression plasmids with promoters of different strength. Plasmid pHRS-Sce1 expresses I-SceI under the weaker $tetO_2$ promoter (two tetO boxes) whereas pHRS-Sce2 expresses I-SceI under the stronger $tetO_7$ (seven tetO boxes).

Replacing HIS3 with KanMX in the genome of EBY100

Most yeast-display projects use the EBY100 strain of yeast first developed by Boder and Wittrup[14]. It was derived from BJ5465 (**a** ura3-52 trp1 leu2Δ1 his3Δ200 pep4::HIS3 prb1Δ1.6R can1 GAL)[15]. EBY100 and its derivatives utilize a HIS3 insertion into the PEP4 locus in yeast chromosome XVI to give a functional knock-out of protease activity[15]. However, for our purposes, this is a bit of a waste of a perfectly good marker, as selectable markers like HIS3 are the primary means by which yeast geneticists can ensure that their designed plasmids are present in their strain constructions[16]. However, the number of auxotrophic markers that work well (low false positive background) are limited. The four genes HIS3, LEU2, TRP1, and URA3 are robust plasmid markers, and selection is readily done using yeast dropout media[17]. In our genetic system, URA3 is used to maintain the AGA1 gene under control of the GAL promoter in the chromosome. Since this integrative plasmid has a certain frequency of splicing itself back out of the genome, the URA3 marker was necessary to ensure yeast display function. TRP1 and LEU2 were already being used to select for the yeast display plasmid and cassette plasmid, respectively. Without HIS3 available, we would have had to utilize a sub-optimal alternative for the I-SceI plasmid. To free up HIS3, we replaced the HIS3 in PEP4 with a KanMX module, which confers resistance to the G418 antibiotic[18], effectively replacing the selection for growth on His- media with an antibiotic resistance selection. Thus the PEP4 locus would remain disrupted, but we would be able to use the HIS3 gene as a selectable marker for the I-SceI plasmid.

The genomic sequences flanking the PEP4 locus were obtained from the Saccharomyces Genome Database and used to design PCR primers. PCR was used to amplify the KanMX module with flanking homology to the PEP4 locus (Figure 3-3). This cassette was transformed into EBY100 yeast., followed by homologous recombination, would result in complete removal of HIS3 and any residual PEP4, and insertion of KanMX which could be selected.

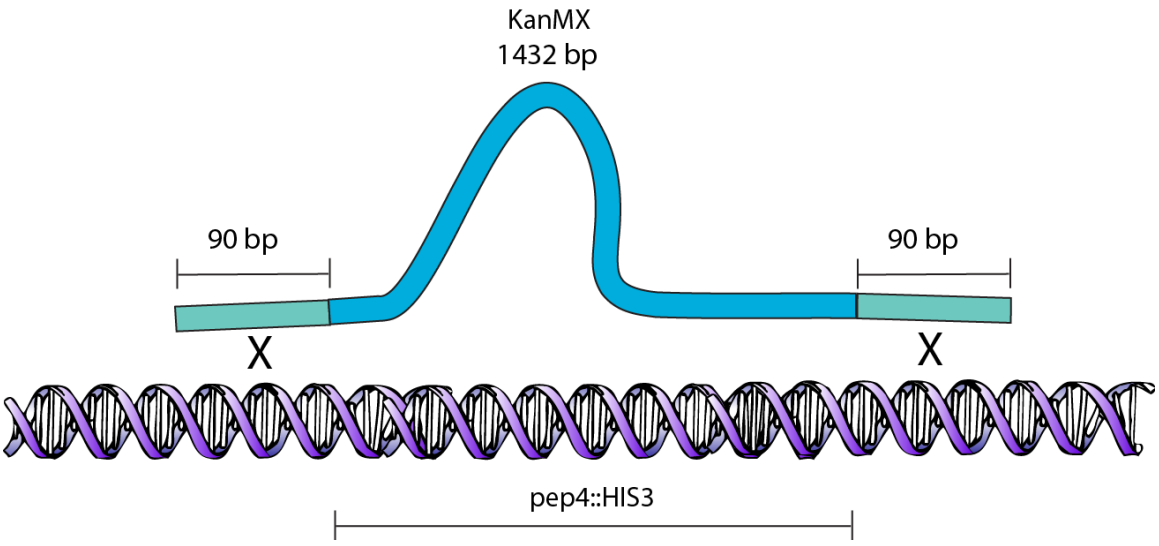


Figure 3-3. Chromosomal replacement of pep4::HIS3 by homologous recombination. Yeast chromosome XVI in strain EBY100 is shown in purple. An insertion cassette was created by PCR, containing the geneticin-resistance marker KanMX (blue) and flanking homology regions (teal). Recombination removes pep4::HIS3, maintaining the functional knock-out of pep4 while also freeing up the HIS3 marker to be used in the I-SceI plasmid.

Transformants were plated on G418 plates (Figure 3-4a). Compared to the negative control, the many transformants clearly acquired G418 resistance. Individual colonies were taken from the G418 plates and then patched onto SC(HU-) and SC(U-) plates to check for inability to grow on H-. If indeed the pep4::HIS3 was replaced by KanMX, then the successful transformants should *not* grow on SC(HU-) plates. Such colonies were found (Figure 3-4b), indicating that it possessed G418 resistance but no longer could grow on His- media. This strain was named ysMDM100.

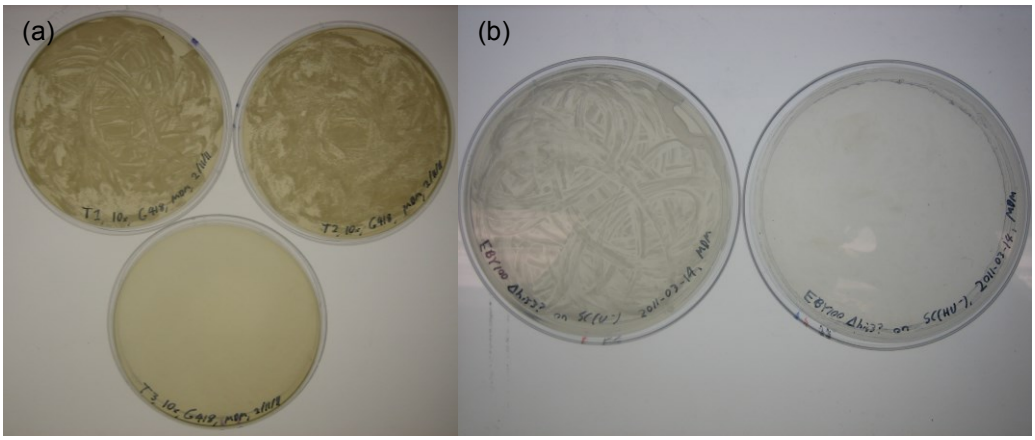


Figure 3-4. Gain of KanMX, loss of HIS3 marker in yeast transformants.
a, G418 antibiotic plates. Growth is observed on the top two plates, which were spread with EBY100 cells transformed with the KanMX insert. No colonies are seen on the bottom plate, which was spread with control EBY100 cells transformed with no DNA. Individual colonies were taken from G418 dilution plates and plated on SC(U-) and SC(HU-). **b,** an example of one such colony, where growth is observed on SC(U-), but not SC(HU-), indicating loss of the HIS3.

Plasmids containing mutagenic cassettes flanked by I-SceI cut sites

To test the ability of the *in vivo* mutagenesis to effect mutagenesis, we constructed cassette plasmids that contained either the functional HA4 BC loop sequence or the placeholder BC sequence. The mutagenic cassettes (Figure 3-5) were constructed by PCR. We chose to make the homology regions 60 bp in length based on previous work that had thoroughly characterized the relationship between the length of flanking homology regions and the recombination efficiency of cassette insertion[21]. 40 bp of homology on both sides was required for high recombination efficiency (> 90%), and diminishing returns were observed for homology > 60 bp.

Figure 3-5. Cassette designed to target a monobody loop for mutagenesis. The mutagenic sequence (green) is flanked on both sides by homology regions (tan), which target the mutagenic sequence to the appropriate location within the monobody gene in the yeast display plasmid. The homology regions are themselves flanked by recognition sites for the I-SceI endonuclease. DSB formation recruits the yeast recombination machinery to the end of the homology region, enhancing the efficiency at which the cassette is targeted to the monobody gene.

These cassettes were cloned into pRS425 plasmids via ligation, resulting in pHRS-Cas2 containing the HA4 BC loop sequence. pRS425 belongs to the YEp class of 2μ vectors[16] which allows for high transformation efficiency and multiple copies of plasmid per cell. This decision was based primarily on the fact that the pRS420 series of plasmids[17] are maintained in their host yeast cells at 10-30 copies per haploid yeast cell[19]. We reasoned that having multiple copies of the cassette plasmid in each cell would mean there would be multiple copies of the cassette that could potentially undergo homologous recombination with the binder DNA on the display plasmid. The higher the copy number of the cassette plasmid, the higher the frequency of homologous recombination would be. Romanini's work showed that high copy cassette plasmids were viable and free of mutation even after induction of I-SceI cutting[20]. Furthermore, it wasn't clear whether the process of generating cassettes would use up or destroy the cassette plasmid involved, so it might be required for additional copies of a given cassette plasmid to be present in order to pass the genetic information onto offspring cells.

Assays for I-SceI induced homologous recombination

ysD2BC2S2, was made by co-transforming pHRS-YD2B (HA4), pHRS-Cas2 (HA4 cassette), and pHRS-Sce2 (strongly promoter I-SceI) into ysMDM100. This strain was a positive control strain because both the yeast display plasmid and the cassette plasmid contained the functional HA4 BC sequence. ysD1BC2S2, was made by co-transforming pHRS-YD1B (HA4 with placeholder BC), pHRS-Cas2 (HA4

cassette), and pHRS-Sce2 (strongly promoter I-SceI) into ysMDM100. This was the experimental strain; the non-functional BC loop in the yeast display plasmid could only be fixed by recombination with the BC loop in the cassette plasmid.

The first assay for homologous recombination that we developed was based on a differential PCR reaction (Figure 3-6). Two sets of primers were designed, with one set optimized to prime from the unrecombined (placeholder) BC DNA sequence, and the other set optimized to prime from the recombined (HA4) sequence. The 5' primer in both sets was designed to bind within the BC loop. The 3' primer was designed to bind further downstream, far enough that PCR could only take place using the BC loop in the yeast display plasmid and not the BC loop contained within the mutagenic cassettes also present within the same yeast cell.

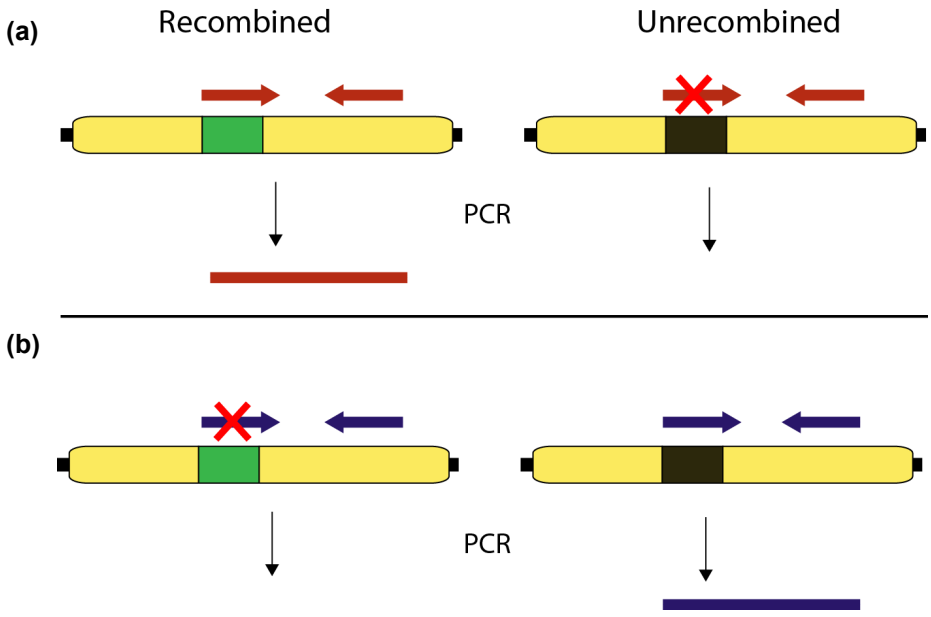


Figure 3-6. A PCR assay to detect homologous recombination of the BC loop. **a,** One set of primers (red) is designed to bind to the functional HA4 BC loop (green). **b)** A different set of primers (dark blue) is designed to bind to the template BC loop (black). The PCR products (colored bands at bottom) can be visualized on an agarose gel.

We first did a control experiment to see if this assay could distinguish between cells containing an unrecombined BC loop and cells containing a recombined HA4 BC loop sequence.

Colony PCR was carried out with cells from both strains ysD2BC2S2 and ysD1BC2S2.

The results are shown in Figure 3-7. As designed, strong, clean PCR products were obtained when the primer set matched their intended target. When the primer sets were used on the alternative BC sequence, the resulting PCR patterns on the gel were visually distinct, with multiple weak bands at incorrect gel distances. In this way, the homologous recombination state of the BC loop can be observing which primer set gives the PCR product at the correct gel distance.

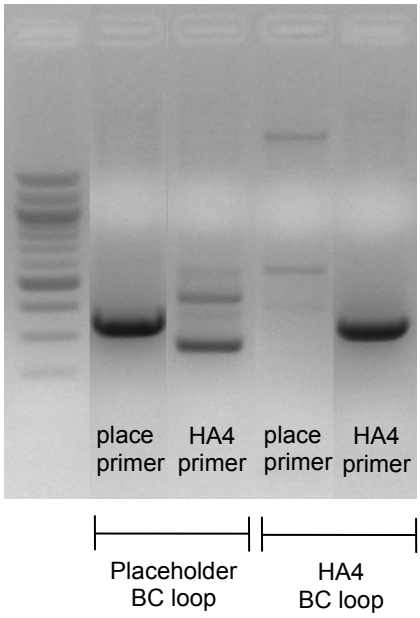


Figure 3-7. PCR assay can distinguish placeholder BC from HA4 BC. In the case of cells being PCRed with the matching primer set (lanes 2 and 5), a single, strong band is produced. In the case of cells being PCRed with the non-matching primer set (lanes 3 and 4), weaker, non-specific bands are observed. "Place primer" denotes reactions where the PCR primers were designed to bind to the placeholder BC loop, and similarly for "HA4 primer".

Doxycycline timecourse for induction of homologous recombination

Having verified the differential PCR assay's ability to distinguish HA4 BC loop sequences from unrecombined placeholder BC sequences, we then tested to see if doxycycline induction of I-SceI expression could lead to conversion of a placeholder BC to HA4 BC *in vivo*. We didn't know *a priori* what length of time would be optimal for the doxycycline induction, so this experiment was performed as a time trial.

The experimental strains had I-SceI expressed under either a weak $TetO_2$ (ysD1BC2S1) promoter or a strong $TetO_7$ promoter (ysD1BC2S2). The positive control strain ysD2BC2S2 was run alongside as a PCR control. Doxycycline was added to cell cultures at OD_{600} = 2 to a final concentration of 1 µg/mL. The cells were kept shaking up to seven days with the doxycycline. Cells were removed at various timepoints, and colony PCR with the differential primers was used to produce the diagnostic PCR products (Figure 3-8). Before addition of doxycycline, only the positive control gave a specific PCR product. After 3 days, the cells with I-SceI expressed under the strong promoter had a distinct PCR product. By about 4 days, the cells with the weak promoter have a noticeable band, and after 7 days, both strains yield a well-defined PCR product.

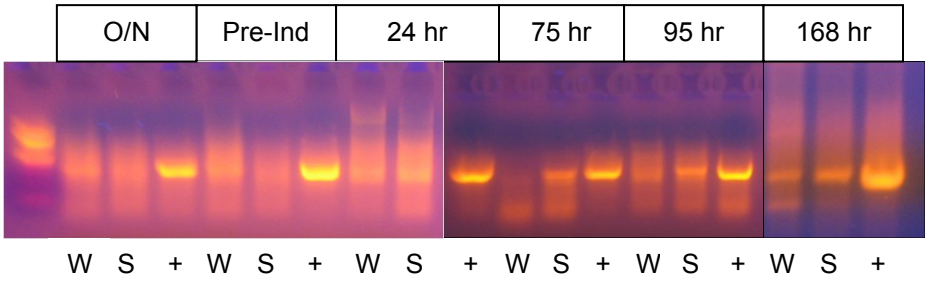


Figure 3-8. Increased homologous recombination with longer doxycycline induction. Duration of cell exposure to doxycycline is given along the top, O/N are cells from the overnight culture prior to addition of doxycycline, and Pre-Ind are cells taken immediately prior to the addition of doxycycline. W = ysD1BC2S1 (I-SceI expressed by weak $TetO_2$ promoter), S = ysD1BC2S2 (I-SceI expressed by strong $TetO_2$ promoter), and '+' = ysD2BC2S2 (positive control where BC loop is HA4 to begin with).

Western blot of I-SceI expression

Protein expression of the I-SceI endonuclease under both the $tetO_2$ and $tetO_7$ promoter system was analyzed by western blot. The I-SceI protein was labeled by its N-terminal FLAG tag (Figure 3-9). The following strains were analyzed: (numbers correspond to gel lanes in figure)

3. Haploid, $TetO_2$ promoter, uninduced
4. Haploid, $TetO_2$ promoter, induced 22 hrs with doxycycline
5. Haploid, $TetO_7$ promoter, uninduced
6. Haploid, $TetO_7$ promoter, induced 22 hrs with doxycycline
7. Diploid, $TetO_7$ promoter, induced 22 hrs with doxycycline

Both the diploid $TetO_7$ and the haploid uninduced $TetO_7$ appeared to be producing soluble protein that could be labled with anti-FLAG antibody.

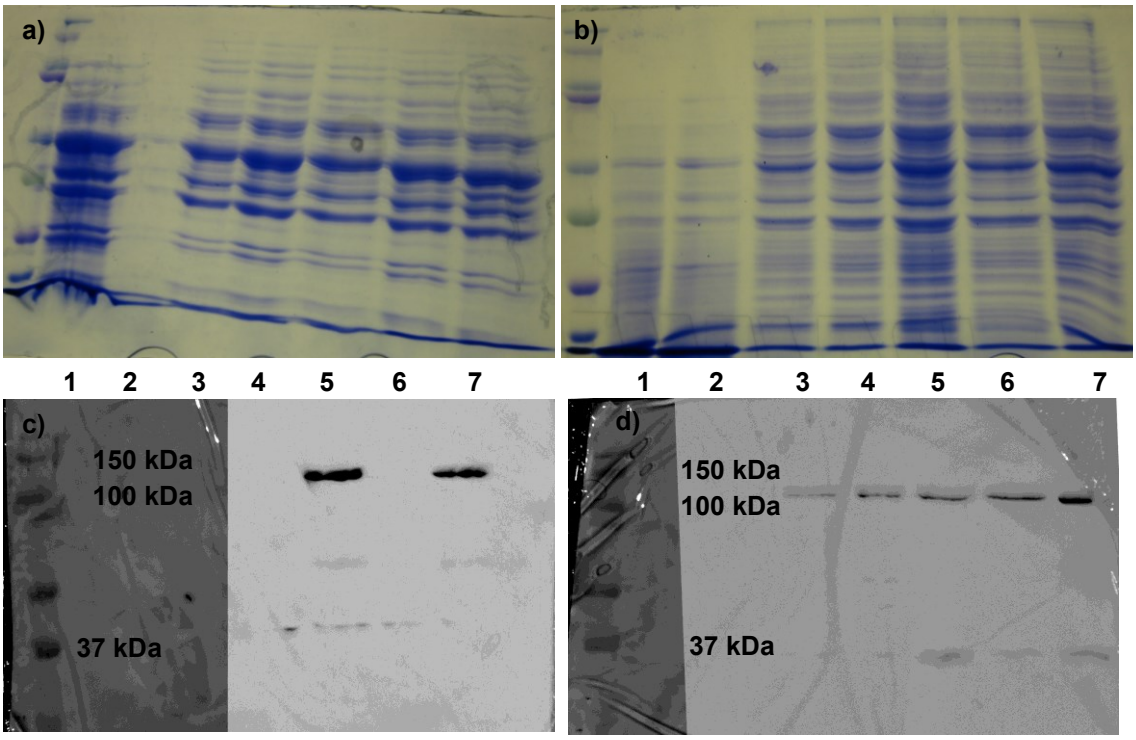


Figure 3-9. Western blot analysis of I-SceI expression under reverse tet transactivator system. **a,** Soluble fraction Coomassie stain. **b,** Pellet fraction Coomassie stain. **c,** Soluble fraction western immunoblot. **d,** Pellet fraction Coomassie immunoblot. The doxy induced diploid shows expression, at both the expected size 29.5 kDa and also at higher bands. The band between 100 and 150 kDa seems to be related to the band below 37 kDa, as they show the same expression pattern in lanes 5 and 7. Furthermore, it seems that all the tet system samples (3-7) have the larger band in the pellet fraction.

Mock libraries to determine recombination efficiency

The assay up to this point has provided evidence of successful recombination at the DNA level. It still remained for us to demonstrate that the recombinants had actually reconstituted functional HA4 which could be expressed on the surface and used in enrichments. ysD1BC2S2 (called “C2” in this experiment) was used as the experimental strain, as it was in the assays described above. For a negative control, a strain was constructed (ysD1BC4S2 or “C4”) which possessed a cassette plasmid that contained the same non-functional placeholder BC that the yeast display plasmid had. Even if homologous recombination were to take place in C4, it would not produce a functional binder.

The overall idea is illustrated in Figure 3-10. C2 and C4 are mixed in different ratios.

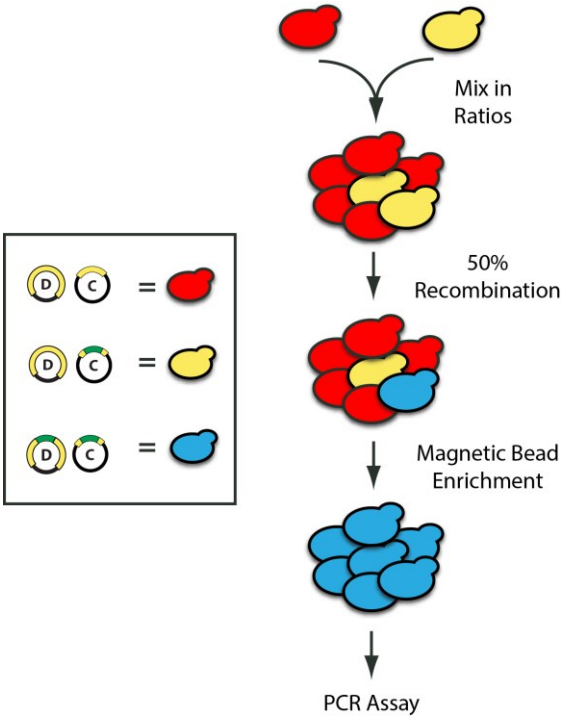


Figure 3-10. Detection and Enrichment of Recombinants from Mock Libraries. Yellow "C2" cells possess a non-functional HA4 in the display plasmid and a fixing loop sequence in the cassette plasmid, and thus has the potential to reconstitute a functional binder. Blue cells are C2 cells that have realized this potential by successfully recombining the fixing cassette into the display plasmid. Red "C4" cells possess the non-functional HA4, but a dummy cassette plasmid, and thus have no ability to reconstitute a binder. By mixing yellow cells with an excess of red cells we can make a "mock" library. In the example shown, C2 and C4 cells are mixed in a proportion of 2:5.

The cells were mixed in ratios ranging from 1:10 C2:C4 to $1:10^4$ C2:C4, in addition to running a 100% C2 sample and a 100% C4 sample, with all cell samples containing a total of 10^8 cells at the time of mixing. The cell mixtures were induced in 2 µg/mL doxycycline for 72 hrs, and then multiple rounds of magnetic bead selection were conducted. After three rounds of enrichment, a PCR assay was done based on restriction digestion of the PCR products (Figure 3-11). The placeholder BC loop contains a ClaI cut site that the HA4 loop does not. The HA4 cassette contains a BsiWI cut site that the template loop does not. Therefore, cutting individually with these two restriction enzymes should give opposite cutting patterns when visualized on a gel.

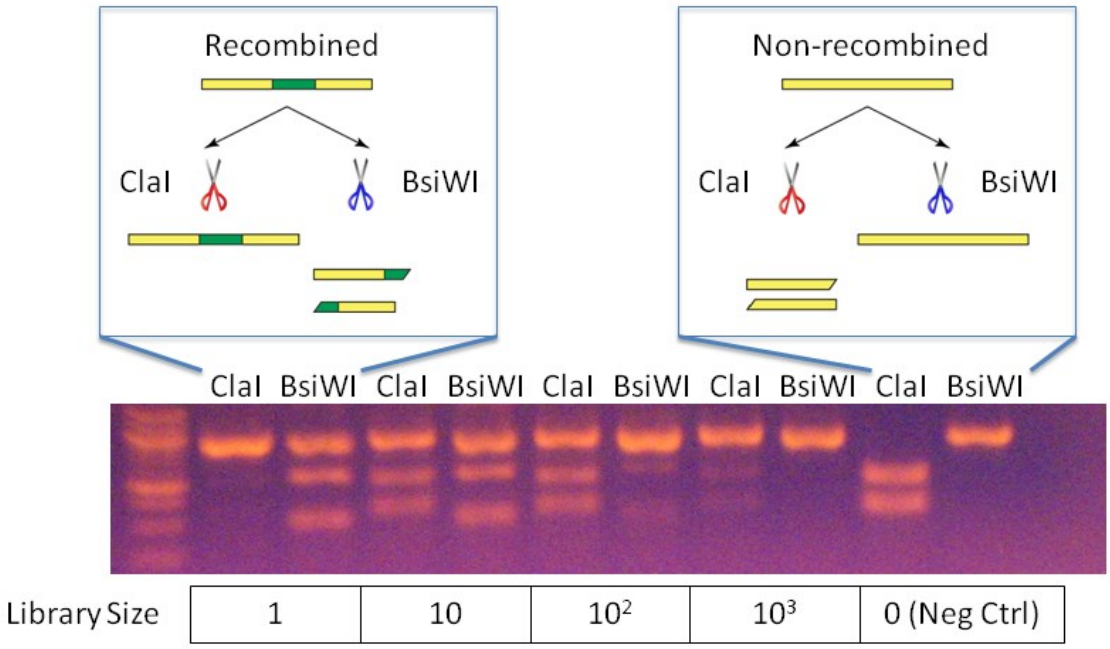


Figure 3-11. Library size determination by recombination and enrichment. DNA coding for the unrecombined BC loop cuts with ClaI but not BsiWI, whereas DNA coding for the recombined BC loop cuts with BsiWI but not ClaI, and thus there should be a mirror flip. Library sizes are shown along the bottom and represent the minimum library size the initial proportion of C2:C4. Thus mock library which started out with $1:10^3$ C2:C4 demonstrated reduced ClaI cutting compared to the control, and thus suggests that we can create libraries at least 10^3 in size.

Indeed, library dilutions up to 10^3 enriched cells missing the ClaI digestion site, which would be consistent with enrichment of cells that had replaced the placeholder BC.

To corroborate these findings with flow cytometry, we labeled the 10^2 dilution mock library with anti-V5 against the yeast display construct (Figure 3-12).

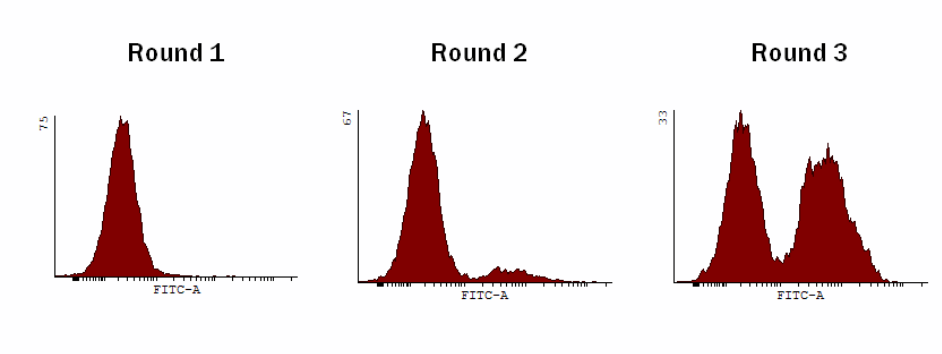


Figure 3-12. Flow cytometry shows enrichment of cells displaying non-truncated monobody on surface. Using 10^2 dilution library. Round # indicates rounds of magnetic bead enrichment. Cells were labeled with FITC-anti-V5.

Mating type switch in ysMDM100 to create mating partner

Since, the EBY100 strain upon which all the yeast display strains are based is of mating type **a**, we needed to create a strain of the opposite type of mating type α, so that we could bring libraries of cassette plasmids together. To do this, we used inducible expression of HO-endonuclease[22] to create a mixed population of the two mating types, which were then separated into individual colonies by plating. The mating type of individual colonies was determined by PCR[23]. The results are shown in Figure 3-13.

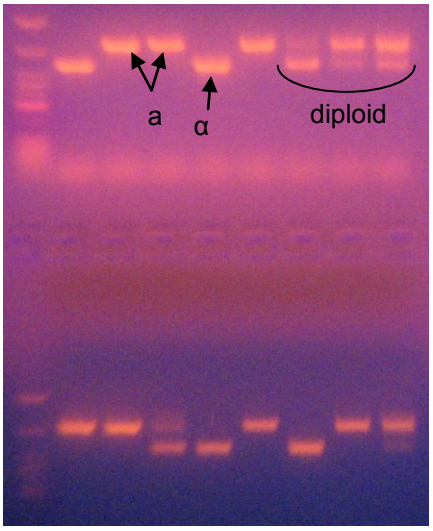


Figure 3-13. Galactose induction of HO endonuclease results in a, α, and diploid cells. The PCR products of 16 different colonies from the same plate are shown here. All three possibilities are represented, with MAT **a** cells producing a band above 500, MAT α cells producing a band at below 500, and diploids producing both bands.

After obtaining the separate mating type strains, the plasmid carrying the HO endonuclease was removed using a 5-FAA TRP1 counter-selection[24].

3.4 Materials and Methods

Construction of yeast display plasmids

BC loops were constructed from synthetic oligos co-transformed into yeast with cut plasmid pGalR38A for gap-repair formation of the desired plasmid. To make pHRS-YD2B, oligos MDM0066 and MDM0067 were extended by PCR, and the product was gel purified and co-transformed into EBY100 yeast with ClaI-digested pGalR38A. Colony growth on SC(UT-) plates indicated gap-repair, which was confirmed by sequencing using primers MDM0035 and MDM0032. To make pHRS-YD1B, pHRS-YD2B was transformed into dam^-/dcm^- competent *E. coli* (New England BioLabs, Cat. # C2925), prepped and double digested with BclI and BsiWI (2.4 µg plasmid, 5 µL NEB buffer #3, 4 µL of BsiWI, 1 µL BclI, 50 µL total, 50 °C overnight). An insert BC sequence was produced by PCR extension (10 µL of 10x Thermopol buffer, 1 µL of dNTP mix, 1 µL of MDM064A, 1 µL of MDM064C, 86.5 µL of ddH_2O, 0.5 µL of VENT DNA polymerase). The insert was purified by agarose gel and co-transformed into EBY100 yeast with the double-digested pHRS-YD2B. Colony growth on SC(UT-) plates indicated gap-repair, which was confirmed by sequencing using MDM0035 and MDM0032.

Construction of cassette plasmids

To make pHRS-Cas2, a DNA insert containing the BC loop sequence of HA4 was created by PCR: 10 µL 10x Thermopol buffer, 1 µL of dNTP mix, 0.5 µL of oligo MDM0075, 0.5 µL of oligo MDM0076, 86.5 µL H_2O, 1 µL of pHFT2-HA4, and 0.5 µL VENT DNA polymerase. The insert was gel extracted, and both insert and plasmid pRS425 were double-digested with SalI and BamHI in NEB buffer 3 + BSA at 37 °C overnight. Cut plasmid and cut insert were purified by agarose gel. 5.4 ng of cut pRS425, 183 ng of cut insert, 2 µL of 10x ligase buffer, and 1 µL DNA T7 ligase were combined in a 20 µL total and incubated at 16 °C for 10 hrs to give pHRS-Cas2. Sequencing was confirmed using primer VM129.

Replacing HIS3 in EBY100 with KanMX

The Saccharomyces Genome Database (www.yeastgenome.org), was used to retrieve the upstream and downstream sequences of PEP4 locus.

Using primers MDM0120 and MDM0121, I copied the KanMX marker from plasmid HO-poly-KanMX4-HO[25]. The reaction mixture was: 1 µL oligo MDM0120, 1 µL oligo MDM0121, 1 µL dNTP mix, 10 µL of 10x Thermopol buffer, 86 µL ddH_2O, 0.5 µL of plasmid HO-poly-KanMX4-HO, 0.5 µL of VENT DNA polymerase. 1 µL of this PCR reaction was transferred directly into a second reaction mix for PCR extension: 1 µL MDM0130, 1 µL MDM0131, 1 µL of dNTP mix, 10 µL 10x Thermpol buffer, 85.5 µL of ddH_2O, 0.5 µL of VENT DNA polymerase; program VENTLONG with 1 m 40 s extension time and Tm = 52 °C. The PCR product was gel purified, and 1 µg was pelleted by ethanol precipitation and transformed into yeast strain EBY100. Transformants were plated on G418 (geneticin) media plates to select for successful insertion of KanMX. Colonies that grew on geneticin were then patched onto both SC(HU-) and SC(U-). If a strain grew on SC(U-) and not SC(HU-), it was deemed to have the HIS3 removed. Further evaluation was made by colony PCR of the yeast strains using the yeast colony PCR protocol (see Section 2.X.X) with 5 µL of 10x Thermopol buffer, 0.5 µL of 100 µM MDM0140, 0.5 µL of 100 µM MDM0141, 0.5 µL of dNTP mix, 0.5 µL of cell lysis mix, 42.5 µL of ddH_2O, and 0.5 µL of VENT DNA polymerase. This reaction was carried out with a thermocycler program of 52 °C annealing temperature and 2 min extension time. Products were run on a 1.0% agarose gel at 120 V.

Construction of I-SceI plasmid

An insert containing the I-SceI ORF was made by PCR: 1 µL of oligo MDM0091, 1 µL of oligo MDM0092, 1 µL of plasmid pDR001 containing the I-SceI ORF, 10 µL of 10x thermopol buffer, 1 µL of dNTP's, 85.5 µL ddH_2O, and 0.5 µL of VENT DNA polymerase, with 50 °C annealing temperature and 1 min extension time per cycle (VENTLONG). The PCR product was purified from a 2.0% agarose gel, run at 120 V for 20 min.

The PCR product, along with plasmids pCM251 and pCM252 were double digested: 5 µL NEB4 buffer, 5 µL BSA, 35 µL of DNA to be digested, 2.5 µL of SbfI, and 2.5 µL of PmeI, and incubated overnight in a 37

°C water bath. The digested DNA was gel purified, and then ligated: 2 µL of 10x ligation buffer, 4 ng of plasmid (either pCM251 or pCM252), 130 ng of I-SceI insert, 1 µL of T4 DNA ligase, to a total of 20 µL, incubated at 16 °C for 10 hrs. The ligations were then pelleted with EtOH precipitation and transformed into competent *E. coli* via electroporation. Transformants were plated on LB + ampicillin plates. The intermediate products pSceα1(from pCM251) and pSceα2 (from pCM252) were purified from these transformants via Qiagen mini-prep kit. pSceα1 and pSceα2 were digested with EcoRV: 5 µL of NEB3 buffer, 5 µL BSA, 39 µL of plasmid (2 µg total), 1 µL of EcoRV, incubated in a 37 °C water bath for at least 4 hrs.

A HIS3 insert was created by PCR of the HIS3 from pRS423: 1 µL of MDM0150, 1 µL of MDM0151, 1 µL of dNTP mix, 1 µL of pRS423, 10 µL of 10x thermopol buffer, 85.5 µL ddH_2O, 0.5 µL VENT DNA polymerase. This HIS3 insert and the EcoRV-digested pSceα1/2 plasmids were gel purified by Qiagen gel extraction kit, pelleted and co-transformed into the desired yeast strain. Gap-repair resulted in the desired pHRS-Sce plasmids.

Time-course trial of doxycycline induction

ysD1BC2S1, ysD1BC2S2, and ysD2BC2S1 were inoculated from patches into 5 mL cultures of SC(HUTL-) in a 30 °C shaker overnight. The following day, took 1 mL of each culture (OD between 0.5 and 1.4) for my "O/N" colony PCR sample, and transferred an additional 1 mL from each culture into 50 mL SC(HUTL-) in flasks, and put shaking at 30 °C for an additional night. The next day, I took 10^8 cells from each flask for "Pre-Induction" sample, and then added 1 mg/mL doxycycline to a final concentration of 1 µg/mL in the flasks, and continued shaking. A sample of 10^8 cells was taken roughly every 24 hrs for up to 7 days. The samples were immediately prepped using glass beads (see Chapter 2, Materials and Methods) and used in PCR: 10 µL GoTaq Buffer, 0.5 µL MDM0032, 0.5 µL of MDM0173, 0.5 µL of dNTP mix, 38 µL of glass bead prep of yeast cells, 0.5 µL of GoTaq DNA Polymerase; thermocycled with 52 °C annealing temperature.

I-SceI expression western blots

Made fresh SDS-PAGE gels

Made fresh transfer buffer (1 L of 25 mM Tris, 150 mM NaCl, 0.05% Tween20, pH 7.5) using 3.036 g Trizma base, 8.766 NaCl, 500 mL of Tween20 for 0.05%.

Empty microfuge tubes were weighed, and then 1 mL of each induction culture was added, spun 14 krpm for 30 sec, and had supernatant removed with pipette. Cell pellets were determined by weighing again.

Sample	Weight tube (g)	Weight pellet (mg)	Y-PER (µL)
1	1.0398	18.8	94
2	1.0345	13.5	68
3	1.0419	20.9	104
4	1.0354	14.4	72
5	1.0412	20.2	101
6	1.0369	15.9	79.5
7	1.0347	13.7	68.5

Lysed yeast cells with the indicated amount of yeast protein extraction reagent (Y-PER from Thermo Scientific). Vortexed until pellet fully suspended (at least 10 sec). Put cells on rotator at RT for 20 min. Spun down cells from rotator 14 krpm, 10 min. Took sup, added 6X SDS buffer. Took pellet, added 30 µL H2O, 30 µL 6X SDS buffer. Boiled SDS samples for 5 min. Loaded samples into gel and ran 1 hr at 140 V. Ran gels for pellet and sup, and duplicate gels for blotting and Coomassie staining and Western blot. Standard ladder used was Precision Plus Protein™ Kaleidoscope Standards (Bio-Rad).

Cut a gel sized piece of nitrocellulose out and placed in DI water by laying gently on top. Put the gels in transfer buffer, along with filter paper. Cut a notch (for identification purposes) into the bottom right of the supernatant gel, into the top right for pellet gel. Removed the nitrocellulose from the DI water and placed in transfer buffer. Set up Western apparatus in gel room. Placed 2 pieces of filter paper on

bottom, followed by the nitrocellulose sheet, then the first gel onto the stack, and rolled with glass pipette to get air bubbles out. Trimmed to make sure each layer of the stack was the same shape. Placed 2 more pieces of filter paper. Placed 2nd membrane, and then 2nd gel.

Confer with diagram on top of the Western apparatus for further details. Hooked apparatus up to voltage source, ran at constant A = 40 mAmps for 1 hr.

To label the membranes, removed nitrocellulose membranes (making sure not to touch them with fingers). Placed them in 5% dry milk in transfer buffer. Put on shaker for 1 hr. Added labeled anti-FLAG antibody. Shook for 1 hr. Added Amersham ECL prime reagent (GE Healthcare Life Sciences) following included instructions.

Magnetic bead selections

As described in chapter 2

Some bead samples were kept uncoated to serve as negative selections for binders against streptavidin or the beads themselves. The other bead samples were coated with at least 5 pmol of biotinylated target protein overnight on a rotator at 4 °C.

Mock library enrichments

Strains ysD1BC2S2 and ysD1BC4S2 were inoculated from plate colonies into 30 mL SC(HUTL-) and put shaking at 30 °C. 6 samples were made by mixing these two starting cultures in varying proportions. All 6 samples were mixed such that there were 10^8 total cells in each. A sample of 100% “C2” cells and a sample of 100% “C4” cells were made. 4 samples were made by mixing, with 1:10 C2:C4 in the first sample, 1:100 in the second sample, 1:1000 for the third, and 1:10^4 for the fourth. These represented libraries of hits “C2” that could potentially reconstitute a functional HA4 monobody through homologous recombination, with the C2 hits being rarer and rarer in the higher dilutions. After mixing the cells from the appropriate two starter cultures, the 8 samples were spun down at 2800xg, 10 min, 4 °C. The supernatant was poured off and the cells were resuspended in 10 mL of SC(HUTL-) for a starting OD600

of 1, and transferred to 50 mL sterile Erlenmeyer flasks. Doxycycline was added to a final concentration of 2 µg/mL in each sample. The flasks were then put shaking at 30 °C for 72 hrs.

10^8 cells of the doxy induction were spun down 2800xg, 10 min, 4 °C, and were resuspended in 15 mL of SG(HUTL-) media in a 50 mL conical tube. The galactose inductions were put on a rotator at 23 °C. 8 µL of resuspended magnetic beads ($5x10^6$ beads) were added to microfuge tubes (two tubes for each of the 6 samples), and washed 2x with 1 mL of PBSF. 6 bead tubes were resuspended with 10 µL of 3.8 µM SH2-GST-biotin + 90 µL of PBSF for the positive selection coated beads, and the other 6 were resuspended with just 100 µL of PBSF for the negative selection uncoated beads, and all beads were put on a rotator in the cold room (4 °C) overnight.

The next day $4x10^8$ cells were taken from each galactose culture and spun down 2800xg, 5 min, 4 °C. The sup was discarded and each cell pellet was washed once with cold PBSF. Then they were resuspended in 1 mL cold PBSF and added to the uncoated beads for negative selection. The cells + uncoated beads were put on a rotator at 4 °C for 1 hr. Meanwhile, the beads + SH2-GST-Biotin were put on the magnet and washed twice with 1 mL cold PBSF to remove unbound protein. When the incubation on uncoated beads was finished, the cells + uncoated beads were put on the magnet at 4 °C for 5 min. The supernatant was pulled off with a pipette and added directly to the washed, coated beads. These tubes were then incubated on the cold room rotator for an additional hour. After this, the cells + coated beads were put on the magnet for 10 min. The sup was pulled off and the beads were washed 2x with 1 mL cold PBSF. Then the washed beads were resuspended in 15 mL SC(HUTL-) in 125 mL Erlenmeyer flasks and put in the 30 °C shaker. Once the cells reached an OD600 ≈ 1, then another round of selection would be started with galactose induction. After each round, cells were glycerol stocked and also cell prepped for PCR analysis.

Construction of opposite mating type yeast strains

pGal-HO-TRP1 (AKA pFH800[22]) was a generous gift from the Symington Lab. It was transformed into ysMDM100 by electroporation to create strain ysMDM200.

A colony of ysMDM200 from a plate was inoculated into 5 mL of pre-induction culture (SD minimal, 2% raffinose, 0.1% glucose) and put shaking overnight at 30 °C. The cells were then centrifuged at 2800xg

for 5 min. The supernatant was poured off, the cells were washed with 5 mL of sterile ddH_2O. The cells were spun down another 5 min and the supernatant poured off. Then the cells were resuspended in 5 mL induction culture (SD minimal, 2% galactose, 0.1% glucose), and put shaking at 30 °C for 5 hrs. After this, the cells were washed once with ddH_2O and then dilution plated on YPD.

Individual colonies were prepped by SDS protocol (see Chapter 2 Materials and Methods). PCR reactions were run according to Huxley[23]. Recipe: 20 uL GoTaq Buffer, 1 uL ODR0143, 1 uL ODR0144, 1 uL ODR0145, 1 uL dNTP's, 74.5 uL H_2O, 1 uL VENT, and 0.5 uL lysed cells in SDS prep. Thermocycling was for 30 cycles of 92 °C for 1 min, 58 °C for 2 min, 72 °C for 2 min, and the PCR products were run on an agarose gel (Figure 3-X).

To remove the pGal-HO-TRP1 plasmid (to make room for the TRP1 containing yeast display plasmid), a TRP1 counter selection with 5-FAA was used[24]. 5-FAA containing plates were made in the following way. Glucose (12.5 g), Bacto agar (5 g), and ddH_2O water added to 100 mL were mixed together and then autoclaved. Immediately after autoclaving, I added 2X SD (125 mL), 10X amino acid supplement (25 mL, see below), and 5-FAA (1.25 mL of 10% w/v in EtOH), and then poured 25 mL per plate.
The 10X amino acid supplement was made in the following way. 0.011 g tryptophan, 0.025 g adenine, 0.025 g uracil, 0.025 g methionine, 0.025 g arginine, 0.027 g histidine, 0.031 g tyrosine, 0.059 g lysine, 0.081 g leucine, 0.078 g isoleucine, 0.039 g phenylalanine, 0.150 g valine, 0.085 g glutamate, 0.382 g serine were added to a 500 mL flask with ddH_2O added to 100 mL total volume. This was autoclaved, and then while the solution was cooling (at around 50 °C), 0.06 g aspartic acid and 0.2 g threonine were added to the solution, which was swirled the solutes were fully dissolved. 25 mL of this amino acid mix was added to the above 5-FAA plate recipe.

Individual colonies of MAT **a** and MAT α type were grown in 5 mL of YPD overnight at 30 °C, and 10x and 1000x dilutions were plated (100 µL each) onto the 5-FAA plates. This was done gently because the plates were very soft and fragile.

Colonies that grew on the 5-FAA plate were then replica patched onto SC(U-) and SC(UT-) plates for final confirmation of removal of pGal-HO-TRP1 plasmid.

3.5 Strains, Plasmids, and Oligonucleotides

Table 3-1. Strains used in this study

Name	Details	Source/Reference
EBY100	MAT**a** *ura3-52 trp1 leu2Δ1 his3Δ200 pep4:HIS3 prb1Δ1.6R can1 GAL* yeast display Aga system	
ysMDM100	EBY100 with *pep4:HIS3* replaced with *KanMX*	
ysMDM200	ysMDM100 with pGal-HO-TRP1	
ysD1BC2S1	ysMDM100 with pHRS-YD1B, pHRS-Cas2, and pHRS-Sce1	
ysD1BC2S2	ysMDM100 with pHRS-YD1B, pHRS-Cas2, and pHRS-Sce2	
ysD2BC2S1	ysMDM100 with pHRS-YD2B, pHRS-Cas2, and pHRS-Sce1	

Table 3-2. Plasmids used in this study

Name	Details	Source/Reference
pGalR38A		Chapter 2
pHRS-YD1B	Yeast display plasmid expressing HA4 with placeholder BC loop	
pRS423	2μ *HIS3* pBIISK ori amp^R	
pRS425	2μ *LEU2* pBIISK ori amp^R	ATCC #77106
pHRS-Cas2	pRS425 with HA4 BC loop	

pCM251	*tetO*$_2$	
pCM252	*tetO*$_7$	
pSceα1	*tetO*$_2$	
pSceα2	*tetO*$_7$	
pHRS-Sce1	*tetO*$_2$	
pHRS-Sce2	*tetO*$_7$	
pDR001	I-SceI	
HO-poly-KanMX4-HO	*KanMX*	GenBank: AF324728 / [25]
pGal-HO-TRP1	*TRP1 CEN4 ARS1 GAL10 HO nuclease*	Symington Lab/[22]

Table 3-3. Oligonucleotides used in this study

Name	Sequence
MDM0035	CCGGATCGGACTACTAGC
MDM0032	ACTGTTGTTATCAGATCAGCGG
MDM064A	AGTTGTTGCTGCGACCCCGACTAGCCTGCTGATTAGCTGGGATGCATCGATGACATG
MDM064C	AACCGGGGAGTTACCACCGGTTTCACCGTATGTGATACGGTAATATTAAACACTCAGCC ATGTCATCGATGCATC
MDM0066	AAACTGGAAGTTGTTGCTGCGACCCCGACTAGCCTGCTGATTAGCTGGGATGCATCGA TGACATG
MDM0067	AGTTGTTGCTGCGACCCCGACTAGCCTGCTGATCAGCTGGGATGCGCCGATGTC
MDM0075	TAATAATTCCGTCGAC TAGGGATAACAGGGTAATGTTCCGACCAAACTGGAAG
MDM0076	ATGTTGATAAGGATCC ATTACCCTGTTATCCCTATGGTACAGTGAATTCCTGAAC
MDM0091	TACACATGAG GTTTAAAC CGGATTCTAGAACTAGTATGG

MDM0092	ATAATACTAG CCTGCAGG TTATTTCAGGAAAGTTTCGG
MDM0120	GCCTAGTGACCTAGTATTTAATCCAAATAAAATTCAAACAAAAACCAAAACTAAC TTAGCTTGCCTCGTCC
MDM0121	AAGCTCTCTAGATGGCAGAAAAGGATAGGGCGGAGAAGTAAGAAAAGTTTAGC TTTCGACACTGGATGGC
MDM0130	TCTATTAGAATTCTATAAGAAAAGAAAAAAAAAAAGCCTAGTGACCTAGTATTTAATCC
MDM0131	ATATACTATAGTTTTTTTTATTGTTATCTACTTATAAAAGCTCTCTAGATGGCAG
MDM0140	AGC CTA CCA CGT AAG GGA AG
MDM0141	CTG AGG ACA AGA TGG TTT GTC
MDM0150	TATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCT GAT TGT ACT GAG AGT GCA CC
MDM0151	CTTGCAGGCAAGTGCACAAACAATACTTAAATAAATACTACTCAGTAATAAC ATA TCG TAT GCT GCA GCT
MDM0173	GCGCCGATGTCTTCTTC
VM129	CGCCAGGGTTTTCCCAG

3.6 References

1. Benatuil, L., Perez, J. M., Belk, J. & Hsieh, C.-M. An improved yeast transformation method for the generation of very large human antibody libraries. *Protein Engineering Design & Selection* **23**, 155–159 (2010).

2. Sblattero, D., Lou, J., Marzari, R. & Bradbury, A. In vivo recombination as a tool to generate molecular diversity in phage antibody libraries. *Reviews in Molecular Biotechnology* **74**, 303–315 (2001).

3. Sblattero, D. & Bradbury, A. Exploiting recombination in single bacteria to make large phage antibody libraries. *Nature biotechnology* **18**, 75–80 (2000).

4. Plessis, A., Perrin, A., Haber, J. E. & Dujon, B. Site-Specific Recombination Determined by I-SceI, a Mitochondrial Group I Intron-Encoded Endonuclease Expressed in the Yeast Nucleus. *Genetics* **130**, 451–460 (1992).

5. Pâques, F. & Haber, J. E. Multiple Pathways of Recombination Induced by Double-Strand Breaks in Saccharomyces cerevisiae . *Microbiology and Molecular Biology Reviews* **63** , 349–404 (1999).

6. Orr-Weaver, T. L. & Szostak, J. W. Yeast recombination: the association between double-strand gap repair and crossing-over. *Proceedings of the National Academy of Sciences, USA* **80**, 4417–4421 (1983).

7. Szostak, J. W., Orr-Weaver, T. L., Rothstein, R. J. & Stahl, F. W. The double-strand-break repair model for recombination. *Cell* **33**, 25–35 (1983).

8. Krogh, B. O. & Symington, L. S. RECOMBINATION PROTEINS IN YEAST. *Annual Review of Genetics* **38**, 233–271 (2004).

9. Mazón, G., Mimitou, E. P. & Symington, L. S. SnapShot: Homologous Recombination in DNA Double-Strand Break Repair. *Cell* **142**, 648.e1–648.e2 (2010).

10. Niu, Y., Tenney, K., Li, H. & Gimble, F. S. Engineering Variants of the I-SceI Homing Endonuclease with Strand-specific and Site-specific DNA-nicking Activity. *Journal of Molecular Biology* **382**, 188–202 (2008).

11. Gossen, M. *et al.* Transcriptional activation by tetracyclines in mammalian cells. *Science* **268**, 1766–1769 (1995).

12. Belli, G., Gari, E., Piedrafita, L., Aldea, M. & Herrero, E. An activator/repressor dual system allows tight tetracycline-regulated gene expression in budding yeast. *Nucleic Acids Research* **26**, 942–947 (1998).

13. Vaze, M. B. *et al.* Recovery from Checkpoint-Mediated Arrest after Repair of a Double-Strand Break Requires Srs2 Helicase. *Molecular Cell* **10**, 373–385 (2002).

14. Boder, E. T. & Wittrup, K. D. Yeast surface display for screening combinatorial polypeptide libraries. *Nature biotechnology* **15**, 553–557 (1997).

15. Jones, E. W. [31] Tackling the protease problem in Saccharomyces cerevisiae. *Guide to Yeast Genetics and Molecular Biology* **Volume 194**, 428–453 (1991).

16. Sherman, F. An introduction to the genetics and molecular biology of the yeast Saccharomyces cerevisiae. *The Encyclopedia of Molecular Biology and Molecular* **6**, 302–325 (1997).

17. Sikorski, R. S. & Hieter, P. A System of Shuttle Vectors and Yeast Host Strains Designed for Efficient Manipulation of DNA in Saccharomyces cerevisiae. *Genetics* **122**, 19–27 (1989).

18. Wach, A., Brachat, A., Pöhlmann, R. & Philippsen, P. New heterologous modules for classical or PCR-based gene disruptions in Saccharomyces cerevisiae. *Yeast* **10**, 1793–1808 (1994).

19. Christianson, T. W., Sikorski, R. S., Dante, M., Shero, J. H. & Hieter, P. Multifunctional yeast high-copy-number shuttle vectors. *Gene* **110**, 119–122 (1992).

20. Romanini, D. W., Peralta-Yahya, P., Mondol, V. & Cornish, V. W. A Heritable Recombination System for Synthetic Darwinian Evolution in Yeast. *ACS Synthetic Biology* 121024121138004 (2012).doi:10.1021/sb3000904

21. Hua, S., Qiu, M., Chan, E., Zhu, L. & Luo, Y. Minimum Length of Sequence Homology Required for in Vivo Cloning by Homologous Recombination in Yeast. *Plasmid* **38**, 91–96 (1997).

22. Nickoloff, J. A., Singer, J. D., Hoekstra, M. F. & Heffron, F. Double-strand breaks stimulate alternative mechanisms of recombination repair. *Journal of Molecular Biology* **207**, 527–541 (1989).

23. HUXLEY, C., GREEN, E. D. & DUNHAM, I. RAPID ASSESSMENT OF SACCHAROMYCES-CEREVISIAE MATING TYPE BY PCR. *TRENDS IN GENETICS* **6**, 236 (1990).

24. Toyn, J. H., Gunyuzlu, P. L., Hunter White, W., Thompson, L. A. & Hollis, G. F. A counterselection for the tryptophan pathway in yeast: 5-fluoroanthranilic acid resistance. *Yeast* **16**, 553–560 (2000).

25. Voth, W. P., Richards, J. D., Shaw, J. M. & Stillman, D. J. Yeast vectors for integration at locus . *Nucleic Acids Research* **29** , e59–e59 (2001).

Chapter 4

Mating Libraries for the Combination of Loop Diversity

4.1 Abstract

In this chapter, we affinity mature an HA4 binder using the complete heritable mutagenesis system, the individual components of which (having developed the separate components in the previous chapters). We used yeast mating to combine libraries with mutations in either the BC or the FG loops, resulting in a library that had mutations in both loops. Galactose induction was used to initiate homologous recombination, and the resulting yeast were selected with magnetic beads and FACS in order to obtain populations enriched with higher affinity binders. We discuss the different library strategies, and present flow cytometry data that shows how the library populations change after rounds of selection. Higher affinity populations were obtained with two different BC loop diversities.

4.2 Introduction

In the past chapters, we validated the different components that were required to construct a system for mutation and selection *in vivo*, and confirmed their proper functioning. With that done, it was time to begin mating libraries and selecting binders.

Past efforts in affinity maturation have utilized yeast mating to combine diversity. In 2004, Weaver-Feldhaus and colleagues demonstrated that yeast mating could be used to bring together antibody heavy chain (V_H) and light chain (V_k) libraries to create a combinatorial library of Fab fragments[1]. In their system, V_H chains were expressed as fusions to the Aga surface proteins, and V_k chains were translated independently and secreted into the culture medium. These V_k chains could successfully form disulfide linkages to the V_H chains, and thus Fab fragments were constituted on the surface of the yeast. In a demonstration of their system, this group PCR-amplified V_H and V_k sequences from a human volunteer immunized with botulinum toxin to create a $3x10^6$ member V_H library and a $5x10^5$ member V_k library in separate yeast strains. These yeast strains were then mated to give $3x10^9$ unique diploids.

Screening this diploid library with magnetic beads and FACS yielded 5 Fab fragments with affinities for botulinum toxin ranging from 0.8 – 2.1 nM.

Following up on this initial work, they then applied this system towards the affinity maturation of high-affinity antibodies against botulinum toxin[2]. Their starting binders were 3 scFv proteins, one which bound a botulinum neurotoxin A1 domain with K_d = 5 nM, and two which bound botulinum subtype B domains with around K_d = 3 nM. In this system, the heavy chains of the starting binders were not varied, but mating of yeast was used to combine a single heavy chain with a library of $5x10^5$ light chains taken from a human immune repertoire. These library members were sorted with 6 rounds of FACS, and the loop sequences of the obtained hits were grafted onto human IgG1 scaffold. The three hits, one for each of the starting binders, were found to have K_d's ranging from 25.1 – 6.59 pM, with an average affinity increase of 37-fold.

Although this work demonstrated the power of mating to quickly affinity mature binders, it has a major limitation for more general application. Light chains and heavy chains are intrinsically separate. If we wanted to mutate two different regions within the same domain (say a single-domain catalytic enzyme), we would not be able to use this system. Even within the field of antibody binders, it might be the case that we want to combinatorially diversify two loops within the same domain, in which case domain shuffling cannot help. A more general solution is needed that can combine libraries that code for diversity in arbitrary locations within the target protein. In our proposed system, the diversity is encoded in cassettes and is targeted to the proper location within the gene of interest by homology. Thus, separate libraries can target the same domain, or different domains of the same protein, or even altogether separate genes on separate chromosomes for applications like enzymatic pathway engineering. All that matters is the sequence information contained within the flanking homology regions.

If the goal of affinity maturation is to use an already known good binder as a starting point to discover a great binder, we wanted to know if it would be a useful to exhaustively search the neighborhood in sequence space around the good binder. Without having additional biophysical information, it is reasonable to operate on the assumption that mutants that are closer in sequence will tend to be more similar in function?

Decades of alanine scanning mutagenesis experiments have shown that changing a single amino acid can have a very dramatic effect on function. Usually, these experiments aim to impair function. But if these single mutations can significantly impair function, then that implies that the reverse mutation significantly improves function. In affinity maturation, we have our starting sequence, corresponding to good function. We don't want to throw away the good residues, but in general we won't know which ones are good and which ones are bad. So ideally, we would want some way of systematically working through all the mutants in a conservative manner, keeping most of the protein the same.

Instead of leaving it up to chance, we can use the low cost of oligo synthesis to exhaustively search every loop that is similar to the existing loop. How do we define "similar?" The Hamming distance concept, which finds applicability in computer science and information theory, can also be applied to protein and DNA sequences[3]. This seems a reasonable place to start. It is reasonable to hope that the set of all amino acid sequences with H = 1 compared to our starting loop will have many members that have similar properties. Certainly, certain key residues, when mutated to any of the other 19, might abolish binding function. But if we wanted to do an exhaustive search of a certain set of loop sequences that could fit within a limited library size and that had a reasonable chance of containing functional hits, a reasonable strategy might be to search every loop sequence with H < X, X chosen so that the total theoretical complexity can fit within the library size searchable by the available experimental methods.

For example, for a library (set) of all BC loop sequences with H ≤ 2 compared to the HA4 BC sequence (8 residues long), the theoretical diversity would be:

$$\binom{8}{2} \times 48^2 = 6.45 \times 10^4$$

We have a 48^2 term instead of a 20^2 term because we are using NNB codons, where B = C,G, or T nucleotides, thus there are 4x4x3 = 48 different codons, and only 1/48 ways to code for Tryptophan, for example.

Table 4-1. Theoretical "Hamming 2" (H ≤ 2) library complexities for different types of diversity

	NNB codons	NNY codons
8 residue loop	6.45×10^4	2.87×10^4
10 residue loop	1.04×10^5	4.61×10^4
13 residue loop	3.59×10^5	1.60×10^5

This type of library design is extremely amenable to the cassette-based loop swapping that our genetic system is designed for.

I also needed to show that strains based on EBY100 could successfully mate. Wittrup's yeast display system[4] relies critically on the use of Aga1p and Aga2p, proteins on the surface of yeast that function in mating between **a** and α cells. The monobody protein is fused to the Aga2p protein, and that allows it to be displayed on the yeast cell surface. However, fusing an exogenous protein to Aga2p might interfere with its natural function. Work by Nobel and coworkers[5] has showed that *aga2* mutations impair mating efficiency in liquid culture to 0.1% compared to wild-type. Mating efficiency on solid medium was unaffected, which would be consistent with the idea that aga2p facilitates (but is not essential for) **a** and α cell mating by holding the cells in proximity to each other[6]. Since our mating is intended to be done on solid medium, these studies were reassuring. However, we needed to characterize mating in our system, anyway, due to its importance in horizontal genetic transfer.

Another important source of genetic diversity are existing libraries. We decided it would be interesting to make a library of loops taken from a library that was already enriched against the binder, and then mate that combine that library with a library of our own design for another loop. In their paper describing the HA4 binder[7], Wojcik and co-workers obtained an intermediate library with an aggregate K_d > 500 nM, before doing affinity maturation to arrive at the HA4 clone. We decided to pull out the FG loop sequences from this intermediate library and mate that with a BC library of our design.

4.3 Results

Fully random NNB and NNY libraries in cassette plasmid

These libraries are the same type of diversity as the mixed-length BC libraries discussed in Chapter 2. The two libraries (NNB and NNY codon diversity) were transformed into cassette plasmids in different mating types. In order to maximize library complexity, each of the 4 sub-libraries (corresponding to a specific codon length of 6, 7, 8, or 9) was transformed separately, and then the successful transformants were combined.

Table 4-2. Fully Random NNB/NNY Library Complexity

MAT a		**MAT α**	
6 NNB codons	1.7×10^6	6 NNY codons	5.5×10^6
7 NNB codons	6.1×10^6	7 NNY codons	7.7×10^6
8 NNB codons	4.4×10^6	8 NNY codons	4.4×10^6
9 NNB codons	7.7×10^6	9 NNY codons	8.5×10^6
Sum	**1.99×10^7**	Sum	**2.61×10^7**

The 4 NNB libraries of different lengths in MAT **a** yeast were combined to make mixed-length library of complexity 1.99×10^7. The 4 NNY libraries in MAT α were combined to make a library of complexity 2.61×10^7.

An enriched library of FG loops

In their 2010 paper, Wojcik and co-workers built an initial binder library by randomizing the BC, DE, and FG loops and then selecting for SH2 binders with phage display[7]. This gave them an intermediate library, which was then loop-shuffled to give HA4. We obtained this intermediate library (which I will simply refer to as "Wojcik Library") from the Koide Lab and cloned the FG loops into the yeast display plasmid via PCR and gap-repair. We sequenced the colony PCR products from this yeast library to get a sense of the type of library diversity. We found that 48.39% of the FG loops in our sample were the HA4 FG loop sequence (high affinity).

Table 4-3. Sequencing of FG loops found in Wojcik library.

FG Sequence	Occurrence	Frequency
WGEDSAGYMFMYS	15	48.39
WGEDXXGDMFMYS	1	3.23
SYSNLNYGYSYYS	2	6.45
SYSNLN*GYCYYS	1	3.23
FKHKYGYAFYYS	1	3.23
RYPYWMWYSDS	1	3.23
SWYDYYGYSYDS	1	3.23
WHYWFYYYS	1	3.23
WSDGVEYMHESS	1	3.23
WWWYSSSSS	1	3.23
WYHGWQYGS	1	3.23
WYYDLNYG*F*YS	1	3.23
YDYEWWQYS	1	3.23
YGYPEAYGS	1	3.23
YSYFGYDS	1	3.23
YYHTSHYDHYS	1	3.23

Mating full NNB/NNY libraries with Wojcik FG library

To test mating, we took yeast cells containing NNY/NNB cassette plasmids (conferring growth in Leu- media) and mated them with yeast cells containing FG libraries in yeast display plasmids (conferring growth in Trp- media). Only diploids resulting from the mating of the two had both plasmids, and could thus grew on (Leu- Trp- media). $4x10^8$ cells of each mating type were mixed together and allowed to mate on YPD plates for 6 hrs. Following this time, the cells were resuspended and dilution plated on selective media (HU- or HUTL-) plates to determine the number of diploids.

In the mating of NNY library (in MAT α) with ysMDM322 (MAT **a**), $4.4x10^8$ cells ($2.2x10^8$ pairs) were estimated to be capable of growth on HU-, whereas $4x10^5$ cells (or 0.18% of total pairs) were estimated to be capable of growth on HUTL- (and thus counted as diploid). In the mating of NNB library (in MAT **a**) with ysMDM422 (MAT α), $7.0x10^8$ cells ($3.5x10^8$ pairs) were estimated to be capable of growth on HU-, whereas $1.8x10^6$ cells (or 0.51% of total pairs) were estimated to be capable of growth on HUTL-.

Thus we found that we could obtain ~ 10^6 diploids by mating. Since the recombination efficiency is more of a bottleneck, this is a very adequate output.

The diploids were induced in doxycycline, and then enriched by one round of magnetic bead selection.

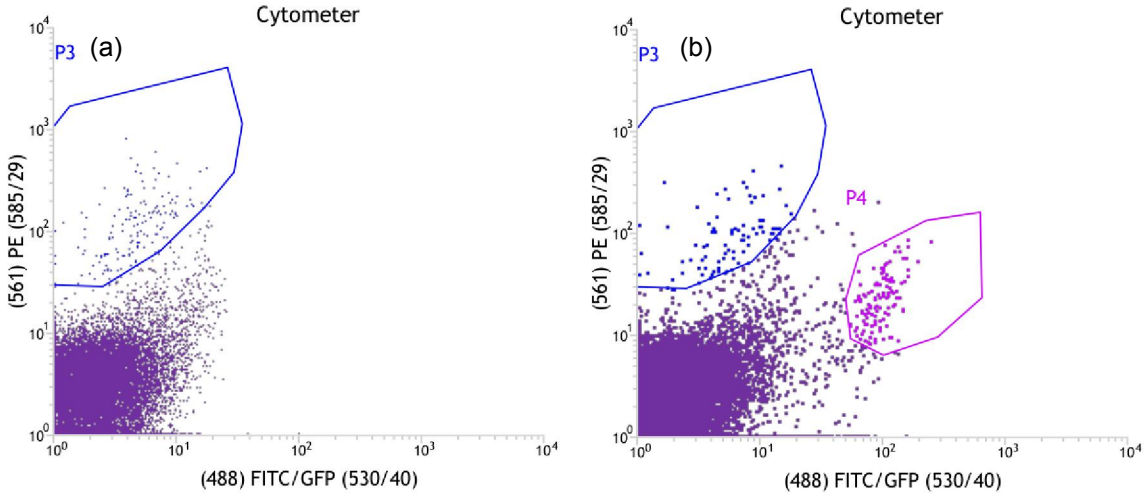


Figure 4-1. Fully random BC x Wojcik FG, after one round magnetic beads.

Both have a significant population of single-positive PE cells (see discussion). **a,** NNY BC x Wojcik FG.

b, NNB BC x Wojcik FG

After switching to the new FITC labeling method, much improved results were observed for an NNB x Wojcik FG library (Figure 4-2).

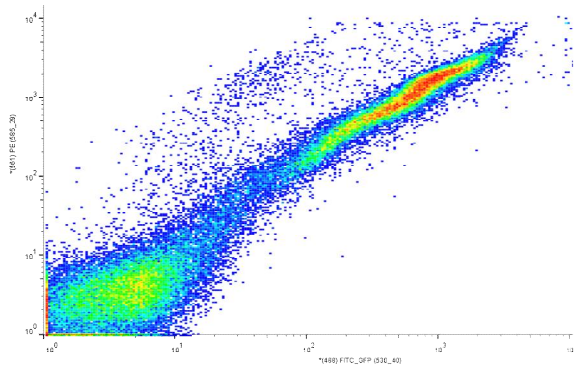


Figure 4-2. NNB x Wojcik FG after 1 round of beads and 1 round of FACS. Higher affinity population is present with lower FITC signal but higher PE signal.

Hamming 2 BC loop library

We decided to randomize residues MSSSSVYY from the HA4 BC loop, which keeps the upstream proline that might force the BC loop into a proper confirmation. This library randomizes the two residues that make direct contact with the SH2 domain.

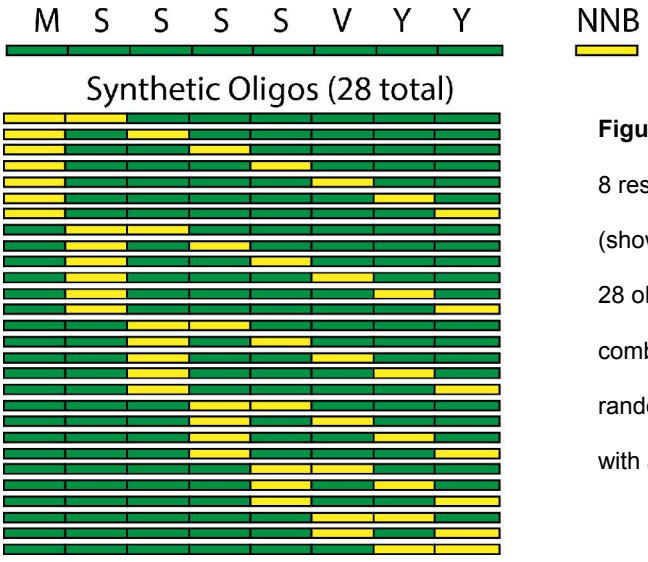


Figure 4-3. Mutational scheme for H2BC library. 8 residues from the HA4 BC loop sequence (shown across the top) were chosen for mutation. 28 oligos were ordered from IDT, with every combination of 2 out of 8 codons being randomized to NNB. This is a low diversity library with a relatively low theoretical complexity.

The transformation of the H2BC library into a cassette plasmid resulted in 6.0×10^6 transformants (strain ysL-CasH2BC). Further transformation with cut yeast display plasmid to give transfer of the library

cassette resulted in $8.6x10^5$ transformants (H2BCgap library), which is still 13X the theoretical library diversity.

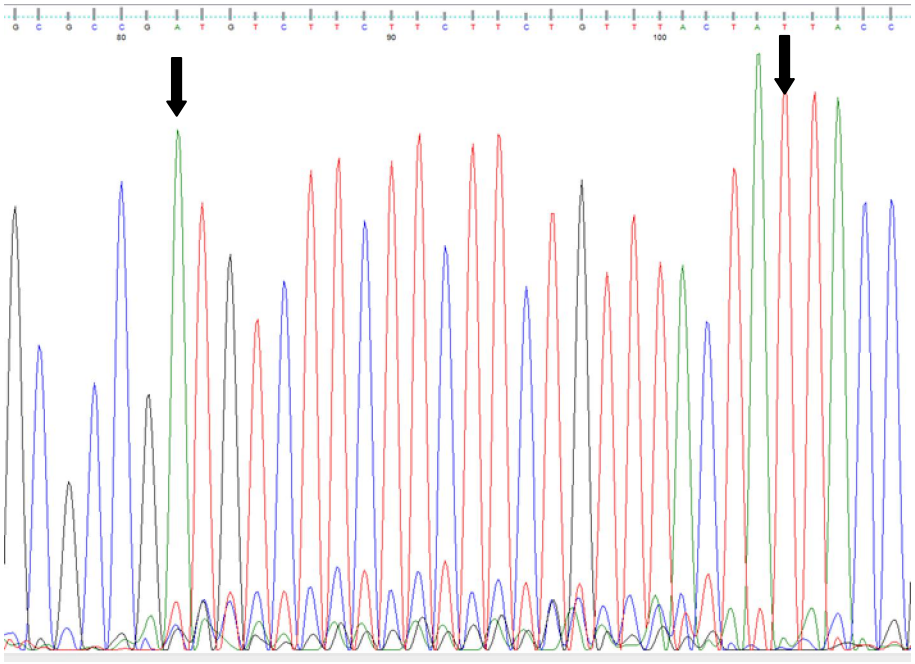


Figure 4-4. Population sequencing of H2BC library. This sequencing data was obtained from the finished PCR insert, prior to transformation and gap-repair into a cassette plasmid. Black arrows mark the first and last nucleotide of the 24 bp of diversified codons. For any specific codon out of the 8, we predict it would be mutated in 7 out of the 28 codons (25% of the time). The other 21 oligos have the wild type nucleotides for that codon. This trace data is reasonable, then, although there appears to be some additional mutation. There is a repeating pattern of absence of A (green) signal every third nucleotide, which is consistent with our NNB library design where B = C, G, or T.

We also constructed a yeast display library of H2BC diversity directly by co-transformation of cassettes with cut yeast display plasmid, with gap-repair giving the desired product, “H2BCvitro”. This library could be compared to the one created by doxycycline induction and homologous recombination.

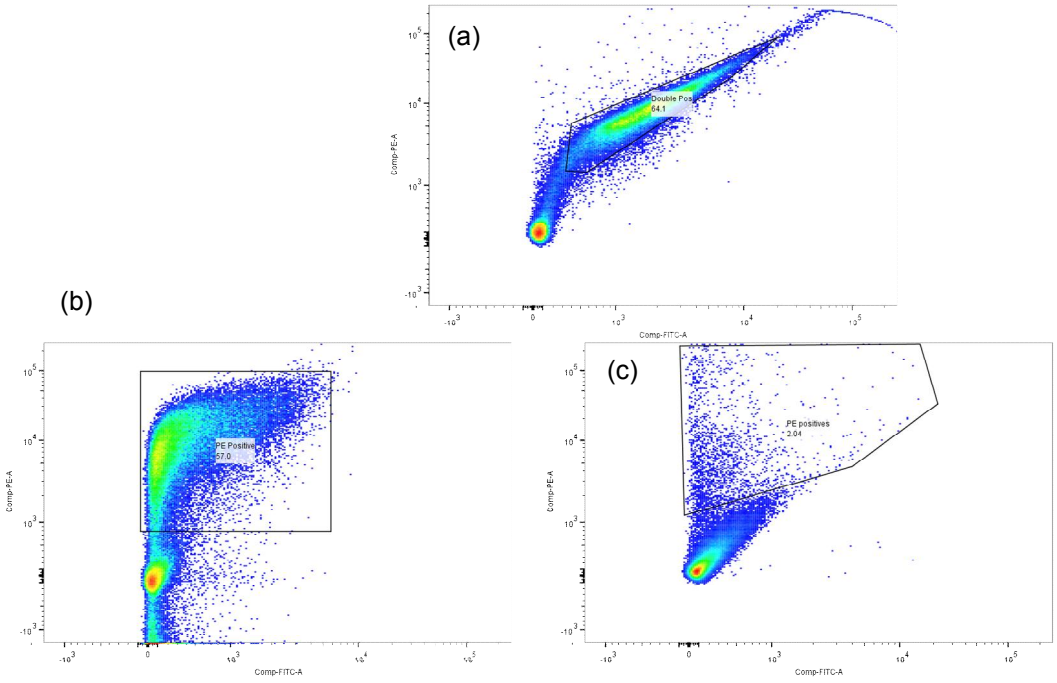


Figure 4-5. FITC labeling scheme inadequate for high cell density. **a,** 10^6 HA4 sample **b,** **$5x10^7$** H2BCvitro, **c, $5x10^7$** NNY x WojFG. Judging from the single-positive populations in b and c, it was clear that the FITC labeling reagent wasn't able to handle the cells

After obtaining the flow cytometry results shown in Figure 4-X, we felt very felt very confident that the current FITC labeling protocol was insufficient. After this point, we switched to a labeling scheme with mouse anti-V5 primary and goat anti-mouse Alexa Fluor 488 secondary.

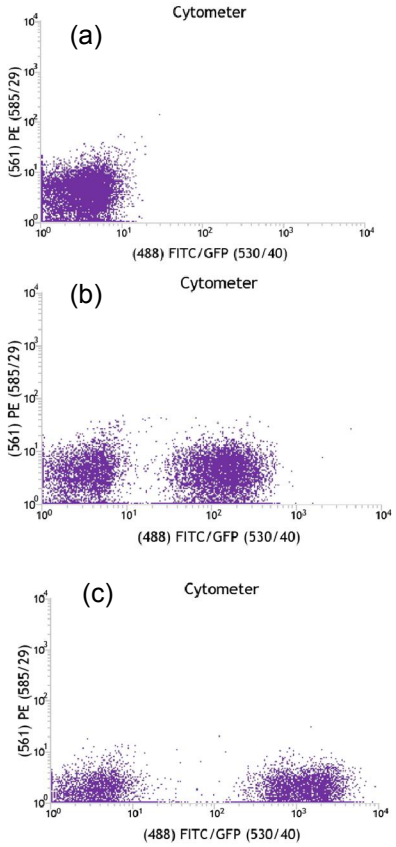


Figure 4-6. An improved method for labeling yeast display. ysEBYHA4 cells labeled with, **a,** no FITC label reagent, **b,** anti-V5-FITC, and **c**, mouse anti-V5 followed by Alexa488-anti-mouse antibody.

In Figure 4-X, we show the FITC signal from ysEBYHA4 populations labeled with no FITC reagent, labeled with the old method (4-Xb), and labeled with the new method (4-Xc). The stark improvement in intensity confirmed the superiority of the primary-secondary labeling method.

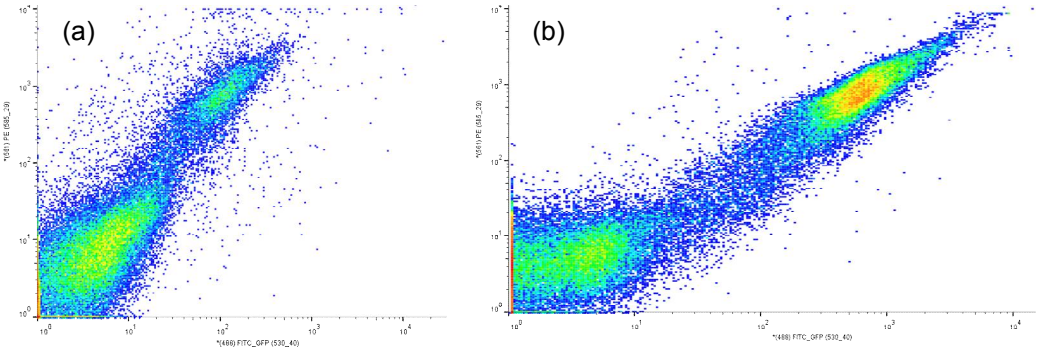


Figure 4-7. H2BCgap after **a,** one round of FACS enrichment, **b,** two rounds of FACS enrichment

Mating H2BC library with Wojcik FG Library

We mated 100x excess of WojFGa with enriched H2BCgap aiming to drive as much of H2BCgap to diploid as possible. Diploids were induced with doxycycline to induce homologous recombination of the BC loop diversity on the cassette plasmid with the FG library contained within the yeast display plasmids.

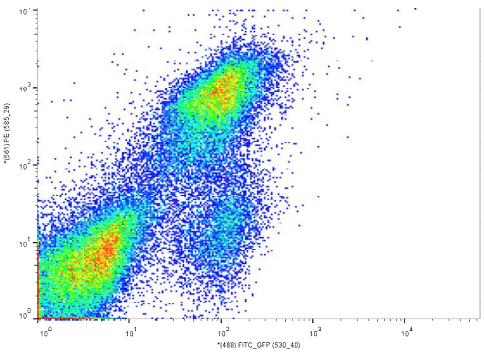


Figure 4-8. H2BCvitro after 1 round of FACS. A large percentage of clones are high-affinity.

Yeast display of peptidase heterozygotes

FY250 (type α) was mated with ysEBYHA4 (type **a)** with 100x excess of FY250. This diploid was then galactose induced and flow cytometry was done to test ability of heterozygote to display yeast on the surface.

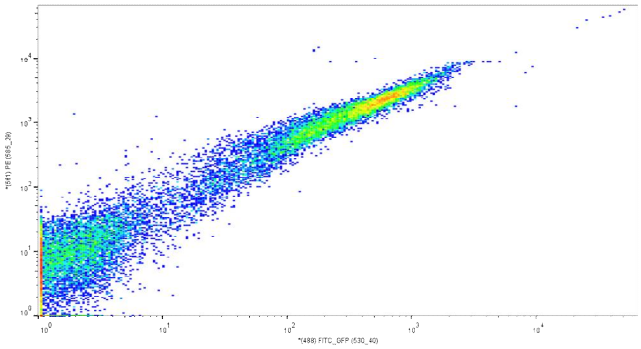


Figure 4-9. Diploids that are heterozygous for yeast display knock-outs can still express monobodies on the cell surface at high copy-number.

4.4 Discussion

The fact that we were doing the FITC labeling sub-optimally in the past explained a lot of times that we've seen a lot of single-positives. On the down-side, it means we actually need to rerun a lot of our libraries. On the plus side, it seems like we have had hits all along in our various libraries.

In this chapter, we begin testing binder libraries that are the result of two loop libraries coming together. For FG loop diversity, we used an intermediate library that ultimately led to the discovery of HA4. Many of these FG loops have affinity for SH2, despite their diverse sequences. This is an interesting opportunity to see if multiple mutational paths can lead to alternative solutions to the binding problem.

For our BC loop library, we have two different library design. First, we used the same completely random, mixed-length libraries that we used in Chapter 2, and mated these to the Wojcik FG loops.

Figure 4-2 showing the NNB x Wojcik FG diploids after 1 round of beads and 1 round of FACS shown is very exciting, and it will be very interesting to sequence the high affinity sequences.

Our second library takes the strategy of a low-diversity library, where no sequence is more than 2 amino acids away from the HA4 BC. I conceptualize this strategy in the following way (Figure 4-X shows the idea for a simplified library).. Take the protein sequence that you are trying to affinity mature, and put it in the center of sequence space, with all other protein sequences of the same residue length arranged around it in concentric circles. On the first circle place all protein sequences that differ by 1 mutational step (Hamming Distance H = 1, following the terminology of Bosley[3]). On the second circle place all the protein sequences that require two mutational steps to reach from your center, and so forth for the rest of the circles. The shells contain more and more sequences and get bigger and bigger until the final shell, which includes all sequences of the same residue length which have no residues in common with your sequence. The Hamming Library mutational strategy is then to figure out the experimental library size that can be searched with our system, and then search as many inner shells are we can with that library size.

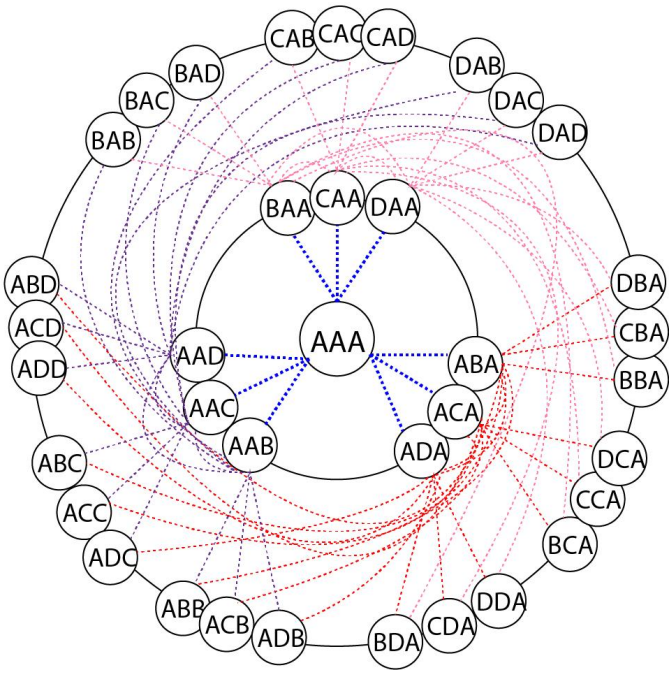


Figure 4-10. Hamming Shells. For a 3-residue long protein, with 4 amino acid diversity. Each concentric circle contains all the nodes of that Hamming distance. The first circle contains all nodes with H= 1 from AAA, the second circle with H=2. There will be as many circles as the residue length of the starting sequence. Third and final circle for this example is not shown for simplicity. Each single mutational step is represented either as a dotted line or as physical contact between nodes.

Our H2BC libraries show many binders with high-affinity, even without enrichment steps. It will be very interesting to sequence these libraries and see what kinds of patterns emerge. Furthermore, with this many functional mutants all slightly different from each other in sequence, it will be interesting to see which Wojcik FG loops are pulled out of the Wojcik FG library.

We are currently in the process of building an H2FG library which will provide another interesting FG library to mate with.

As can be readily seen, the number of possible library mating combinations grows rapidly as we build loop libraries with different types of diversity. That is really one of the great strengths of combining the heritable mutagenesis system with a monobody protein, where the functional loops that we wish to target are well defined. We can create many different loop libraries for the BC and the FG loops, and store them in the fridge. They will be just as valuable for future targets as they are for the current one.

Materials and Methods

Determining mating efficiency

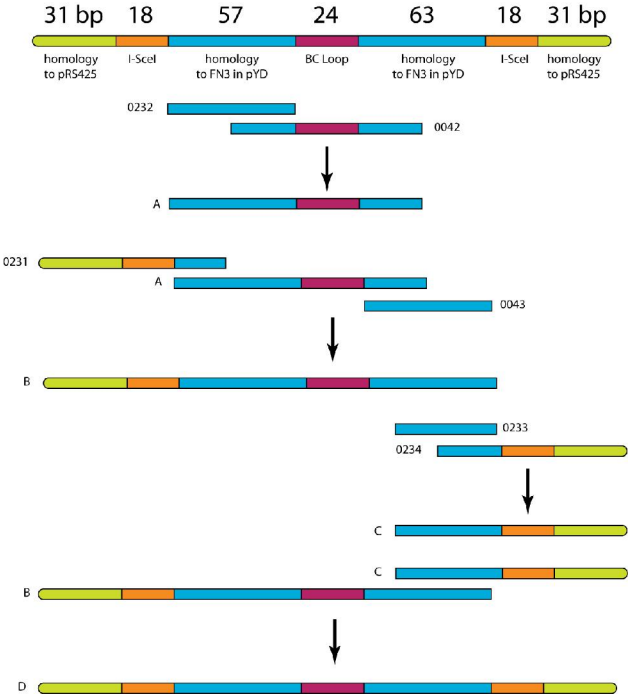
Determined OD600 for cell density of O/N cultures of ysMDM322, ysMDM422, ysL-NNY6789, and ysL-NNB6789. Took 4 x 10^8 cells from each library for 20x coverage of 2x10^7 BC loop complexities. Mixed ysMDM422 with ysL-NNB6789, and ysMDM322 with ysL-NNY6789.

Spun cells at 2800xg, 10 min, 4 °C. Dumped supernatant, resuspended cells in 200 µL YPD and pipetted all of the resuspension into the center of YPD plates without spreading. Put the plates in 30 °C incubator right-side up for 6 hrs. Recovered cells by washing 3x with 1 mL of ddH2O and collecting washes with a pipette. Made dilutions and plated to determine mating efficiency.

Inoculated recovered mated diploids into 50 mL of SC(HUTL-) for propagation and glycerol stocks. 3 days later, counted colonies on the dilution plates to determine # of diploids.

Building fully random BC loop libraries into cassette plasmid

The overall scheme for building the BC mutagenesis cassettes.



The first step of the PCR reaction: 1 µL of MDM0232, 1 µL of dNTP's, 1 µL of 5' oligo (1 of the 8 following: MDM0423-0426 for varying lengths of NNB, MDM0453-0456 for varying lengths of NNY), 20 µL of GoTaq Buffer, 76 µL of H_2Os, and 1 µL of GoTaq polymerase; thermocycled with 10 cycles of 30 sec 95 °C, 30 sec annealing (56 °C), and 1 min extension, with final 10 min extension. Ran PCR products in 2% agarose gel (Figure 4-9) and purified with Qiagen gel extraction kit.

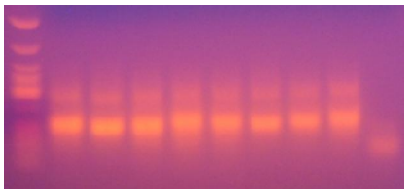


Figure 4-11. First step of random BC cassette synthesis. The varying lengths of the BC loop (6, 7, 8, or 9 codons) can actually be visualized in the slight difference in gel migrations. The bright bottom band is the product, which was excised and purified.

This product was then used in the second step: 1 µL of MDM0231, 0.5 µL of MDM0431, 5 µL (250 ng) PCR product from step 1, 1 µL of dNTP mix, 20 µL of GoTaq Buffer, 71.5 µL of H_2O, and 1 µL of GoTaq DNA polymerase. Thermocycler program was 25 cycles, annealing temperature was 54 °C. A separate reaction was done for each of the 8 loop libraries.

The step 3 PCR reaction created a 3' fragment: 1 µL dNTP mix, 1 µL MDM0234, 0.5 µL MDM0233, 20 µL GoTaq Buffer, 76.5 µL H2O, and 1 µL GoTaq DNA polymerase.

Step 4 was an extension joining of the products directly from step 2 and step 3. Recipe: 40 µL of 5' fragment PCR rxn (1 of the 8 samples from Step 2), 40 µL of 3' fragment PCR rxn (Step 3), 1 µL of dNTP's, and 1 µL of GoTaq DNA Polymerase. Note: neither PCR fragment is purified; they are both coming straight from the previous PCR reactions and thus are still in GoTaq buffer.

This reaction mix underwent 10 cycles of 30 sec melting, 30 sec annealing, 1 min extension. The finished reaction was run on 2% agarose gel at 120 V for 20 min. The band at 250 bp was excised and purified by Qiagen gel extraction kit. This product was used in a final amplification PCR step: 1 µL of MDM231amp, 1 µL of MDM234amp, 1 µL of dNTP mix, 1 µL of purified step 4, 75 µL of H2O, 20 µL of GoTaq buffer, and 1 µL of GoTaq DNA polymerase. The product was gel purified.

Plasmid pRS425 was digested: 5 µL NEB3 buffer, 5 µL BSA, 20 µL of pRS425 maxiprep (15.4 µg), 18 µL of H2O, 1 µL of SalI (20 Units), and 1 µL of BamHI (20 Units). The digestion reaction was incubated in a 37 °C water bath overnight, and then gel purified.

3 µg of this double-cut pRS425 was pelleted with 6.2 – 9.0 µg of one of the 8 BC library diversity cassettes. The 4 NNB diversity cassettes were electroporated into ysMDM320 and the 4 NNY diversity cassettes into ysMDM420. Gap-repair within the yeast resulted in a cassette plasmid encoding a randomized BC loop cassette. Transformants were dilution plated on SC(HUL-) and SC(HU-) to determine transformation efficiency and thus estimate library size. From liquid cultures, the 4 libraries of different length for a give codon type (e.g., NNB) were combined by taking 20x the estimated complexity of each length and mixing into a single flask.

Building Wojcik FG library into yeast display plasmid

The C-terminal end of ^{10}Fn3 was cloned out of Wojcik's plasmid library: 1 µL of 100 µM forward primer MDM242, 1µL of 100 µM reverse primer MDM241, 1 µL of dNTP mix, 20 µL of GoTaq Buffer (5X), 75 µL of H2O, 1 µL of Wojcik pDSB library, 1 µL of GoTaq polymerase; Program: 30 cycles with 58 °C

annealing, 20 sec extension, and 10 min final extension. The bright 195 bp band was excised and purified by gel extraction.

The PCR product was then digested with EcoRI to remove any sequences that non-specifically included the BC loop from Wojcik's library: 34 µL of PCR prep (roughly 2.5 µg), 10 µL H2O, 5 µL 10x EcoRI buffer, 1 µL EcoRI enzyme, 37 °C bath O/N.

pHRS-YD1B was double digested with XhoI and SalI: 5 µL NEB3, 5 µL BSA, 38 µL of pHRS-YD1B, 1 µL XhoI, and 1 µL SalI, incubated in a 37 °C bath O/N. This was gel purified.

3 µg of Wojcik FG insert was then pelleted with 1 µg of double-cut pHRS-YD1B and then contransformed into ysMDM320 and ysMDM420 to make strain ysL-WojFGa and ysL-WojFGα, respectively.

Building H2BC library

1st step: 1 µL of 0232, 1 µL of dNTP's, 1 µL of H2BC mix, 20 µL of GoTaq Buffer, 76 µL of H_2O, 1 µL of GoTaq DNA polymerase, GOTAQSHT program with 10 cycles of 30 sec melting, 30 sec annealing (50 °C), 20 sec. extension, with final 10 min extension. The reaction was run on agarose gel, and the bright product band at 100 bp was excised and gel extracted.

H2BC mix is was created by mixing equal amounts of the 28 individual oligos:

1 Oligo MDMH2BC01 5'-TACGTGATACGGTAVNNVNNAACAGAAGAAGAAGACAT CGGCGCATCCCAGCTGATCAGC-3'
2 Oligo MDMH2BC02 5'-TACGTGATACGGTAVNNGTAVNNAGAAGAAGAAGACAT CGGCGCATCCCAGCTGATCAGC-3'
3 Oligo MDMH2BC03 5'-TACGTGATACGGTAVNNGTAAACVNNAGAAGAAGACAT CGGCGCATCCCAGCTGATCAGC-3'
4 Oligo MDMH2BC04 5'-TACGTGATACGGTAVNNGTAAACAGAVNNAGAAGACAT CGGCGCATCCCAGCTGATCAGC-3'
5 Oligo MDMH2BC05 5'-TACGTGATACGGTAVNNGTAAACAGAAGAVNNAGACAT CGGCGCATCCCAGCTGATCAGC-3'
6 Oligo MDMH2BC06 5'-TACGTGATACGGTAVNNGTAAACAGAAGAAGAVNNCAT CGGCGCATCCCAGCTGATCAGC-3'
7 Oligo MDMH2BC07 5'-TACGTGATACGGTAVNNGTAAACAGAAGAAGAAGAVNN CGGCGCATCCCAGCTGATCAGC-3'

8	Oligo MDMH2BC08	5'-TACGTGATACGGTAATAVNNVNNAGAAGAAGAAGACAT CGGCGCATCCCAGCTGATCAGC-3'
9	Oligo MDMH2BC09	5'-TACGTGATACGGTAATAVNNAACVNNAGAAGAAGACAT CGGCGCATCCCAGCTGATCAGC-3'
10	Oligo MDMH2BC10	5'-TACGTGATACGGTAATAVNNAACAGAVNNAGAAGACAT CGGCGCATCCCAGCTGATCAGC-3'
11	Oligo MDMH2BC11	5'-TACGTGATACGGTAATAVNNAACAGAAGAVNNAGACAT CGGCGCATCCCAGCTGATCAGC-3'
12	Oligo MDMH2BC12	5'-TACGTGATACGGTAATAVNNAACAGAAGAAGAVNNCAT CGGCGCATCCCAGCTGATCAGC-3'
13	Oligo MDMH2BC13	5'-TACGTGATACGGTAATAVNNAACAGAAGAAGAAGAVNN CGGCGCATCCCAGCTGATCAGC-3'
14	Oligo MDMH2BC14	5'-TACGTGATACGGTAATAGTAVNNVNNAGAAGAAGACAT CGGCGCATCCCAGCTGATCAGC-3'
15	Oligo MDMH2BC15	5'-TACGTGATACGGTAATAGTAVNNAGAVNNAGAAGACAT CGGCGCATCCCAGCTGATCAGC-3'
16	Oligo MDMH2BC16	5'-TACGTGATACGGTAATAGTAVNNAGAAGAVNNAGACAT CGGCGCATCCCAGCTGATCAGC-3'
17	Oligo MDMH2BC17	5'-TACGTGATACGGTAATAGTAVNNAGAAGAAGAVNNCAT CGGCGCATCCCAGCTGATCAGC-3'
18	Oligo MDMH2BC18	5'-TACGTGATACGGTAATAGTAVNNAGAAGAAGAAGAVNN CGGCGCATCCCAGCTGATCAGC-3'
19	Oligo MDMH2BC19	5'-TACGTGATACGGTAATAGTAAACVNNVNNAGAAGACAT CGGCGCATCCCAGCTGATCAGC-3'
20	Oligo MDMH2BC20	5'-TACGTGATACGGTAATAGTAAACVNNAGAVNNAGACAT CGGCGCATCCCAGCTGATCAGC-3'
21	Oligo MDMH2BC21	5'-TACGTGATACGGTAATAGTAAACVNNAGAAGAVNNCAT CGGCGCATCCCAGCTGATCAGC-3'
22	Oligo MDMH2BC22	5'-TACGTGATACGGTAATAGTAAACVNNAGAAGAAGAVNN CGGCGCATCCCAGCTGATCAGC-3'
23	Oligo MDMH2BC23	5'-TACGTGATACGGTAATAGTAAACAGAVNNVNNAGACAT CGGCGCATCCCAGCTGATCAGC-3'
24	Oligo MDMH2BC24	5'-TACGTGATACGGTAATAGTAAACAGAVNNAGAVNNCAT CGGCGCATCCCAGCTGATCAGC-3'
25	Oligo MDMH2BC25	5'-TACGTGATACGGTAATAGTAAACAGAVNNAGAAGAVNN CGGCGCATCCCAGCTGATCAGC-3'
26	Oligo MDMH2BC26	5'-TACGTGATACGGTAATAGTAAACAGAAGAVNNVNNCAT CGGCGCATCCCAGCTGATCAGC-3'
27	Oligo MDMH2BC27	5'-TACGTGATACGGTAATAGTAAACAGAAGAVNNAGAVNN CGGCGCATCCCAGCTGATCAGC-3'
28	Oligo MDMH2BC28	5'-TACGTGATACGGTAATAGTAAACAGAAGAAGAVNNVNN CGGCGCATCCCAGCTGATCAGC-3'

Step 2: 1 µL of MDM0231, 0.5 µL of MDM0431, 5 µL PCR product #1, 1 µL of dNTP's

20 µL of GoTaq Buffer, 71.5 µL of H2O, 1 µL of GoTaq polymerase.

Program GOTAQSHT with 54 °C annealing temp, 10 cycles, 10 min final extension. Gel purified band at ≈212 bp.

Step 3: 1 µL dNTP's, 1 µL 0234, 0.5 µL 0233, 20 µL GoTaq Buffer, 76.5 µL H2O, 1 µL GoTaq DNA polymerase; program GOTAQSHT, 10 cycles, 20 sec extension. Gel purified band.

Step 4: 1 µL dNTP's, 44 µL step 3 product (13.86 pmol total), 34 µL step 2 product (13.906 pmol total), 20 µL GoTaq Buffer, 1 µL GoTaq polymerase; Program GOTAQSHT with 10 cycles of 30 sec melting, 30 sec annealing @ 54 °C, 1 min extension. Ran product on 40 mL gel made of 2.5% agarose at 120 V, and excised and gel purified.

Step 5 (a final amplification): 1 µL of 0231amp, 1 µL of 0234amp, 1 µL of dNTP's, 1 µL of product from Step 4, 75 µL of H2O, 20 µL of GoTaq buffer, 1 µL of GoTaq polymerase. Program GOTAQSHT (melt 30 s, anneal 30 sec 54 °C, ext 20 s, 20 cycles). Purified from gel.

1.8 µg of SmaI-cut pRS425 and 12.2 µg of BC library insert were pelleted and co-transformed into the strain ysMDM420 to make strain ysL-CasH2BC.

From ysL-CasH2BC, there were two possible routes to go to insert the library into HA4. The first was to transform cut pHRS-YD2B into ysL-CasH2BC. 3 µg of pHRS-YD2B digested with BclI and BsiWI: 5 µL NEB3 buffer, 23 µL ddH_2O, 20 µL (5.2 µg) pHRS-YD2B, 1 µL (15 U) BclI, 1 µL (10 U) BsiWI; incubated in 50 °C water bath. Double cut plasmid was gel purified and electroporated into ysL-CasH2BC to make ysL-H2BCgap.

For the construction of the H2BCvitro strain, PCR was done with 1 µL of 100 µM MDM0232, 1 µL of dNTP mix, 1 µL of 100 µM MDM0431, 1 µL of 10 µM H2BCmix aas defined above, 20 µL of GoTaq Buffer, 75 µL of H2O, 1 µL of GoTaq DNA polymerase; program GOTAQSHT, 15 cycles of 30 sec melting, 30 sec annealing @ 50 °C, 20 sec extension with final 10 min extension.

Building H2FG library

Just as I have the template BC loop plasmid for the H2BC library, I wanted to build a template FG loop plasmid for the H2FG library: 20 µL of 5x GoTaq (Green), 2 µL of MDM264, 2 µL of MDM266, 2 µL of 10x dilution of MDM265, 1 µL dNTP's, 72 µL H2O, 1 µL of GoTaq DNA polymerase. Purified product from 3% agarose gel.

Digested 2 µg HRS-YD2B for ligation

I will also digest all of my insert (22.7 µg)

Recipe:

5 µL NEB3

5 µL BSA

1 µL SalI

1 µL XhoI

10 µL plasmid or 38 µL insert (FG template cassette)

Purified digested plasmid and insert by gel, and ligated with Quick Ligase Kit (New England BioLabs). 3 µL of plasmid (60 ng) was aliquoted into a microfuge tube .A 3x molar ratio of insert was called for; 5.14 ng insert was added. ddH2O was added to a total of 10 µL. 10 µL of 2x quick ligase buffer was added. 1 µL of quick ligase enzyme was added; mixed rxn thoroughly with a pipette. Incubated rxn at RT on benchtop for 5 min, then put on ice in 2 µL aliquots. Competent TG1 aliquot from -80°C freezer was thawed on ice (5-10 min). Added 50 µL of competent cells to the 2 µL aliquot of ligation rxn; mixed via pipette. Incubated on ice 15 min. Electroporated cells, rescued with 1 mL LB, put shaking at 37 °C. Plated on LB+amp.

Mating H2BC library with Wojcik FG library

Mixed 50 µL Lib13-2 (1.4×10^6) with 5 mL of WojFGa (1.7×10^8 cells)

Mixed 40 µL Lib15-2 (1.3×10^6) with 5 mL of WojFGa (1.7×10^8 cells)

Spun down 10 min at 3500 rpm, then dumped sup

Resuspended cells in 200 µL YPD, then dropped into center of a YPD plate

Put in 30 °C incubator face up at 12:00

Building library strains

pHRS-Sce2 (with I-SceI expressed under the strong $tetO_7$ promoter) was transformed into ysMDM300 and ysMDM400, making ysMDM320 and ysMDM420 respectively.

Doxycycline stocks were made by addiong 3.7 mg to 3.7 mL ddH_2O, mixed by inversion

Filter sterilized through 0.2 micron filter.

Flow cytometry

Cells were induced (total # = 20X library complexity) in selective SG media for 24-36 hrs (Chapter 2). Distributed cells for each sample (generally 10^7 – 10^8 cells) into individual 1.5 mL microfuge tubes. Pelleted cells @ 14,000xg for 30 sec in 1.5 mL microfuge tube. Removed sup with pipette and washed with 1 mL cold PBSF. Scaling for every 10^7 cells in a sample, resuspended the sample in 100 µL PBSF + 1 µL SH2-GST-Biotin + 0.5 µL mouse anti-V5 antibody. Vortexed to break up cell clumps. Put rotating for 30 min – 1 hr, either at RT or 4 °C, depending on desired stringency of selection. Note: all following steps were done on ice. Pelleted cells for 30 sec @ 14,000xg. Washed with ice-cold 1 mL PBSF. Scaling for every 10^7 cells in a sample, resuspended cells in 100 µL of PBSF + 0.5 µL of Streptavidin-PE + 1 µL Alexa488-anti-mouse antibody). Rotated in cold room for 15 min. Pelleted cells for 30 sec @ 14,000xg

Washed with ice-cold 1 mL PBSF. Removed sup and left pellets on ice. Immediately before running cytometry on the sample, resuspend pellet in 250-800 µL of cold PBSF, vortexed briefly.

Labeling

Used 0.5 µL PE in 100 µL (1:200)

Used 1 µL anti-V5-FITC in 100 µL (1:100)

Used 0.5 µL mouse anti-V5 in 100 µL (1:200)

Used 1 µL anti-mouse 488 in 100 µL (1:100)

Testing yeast display in peptidase heterozygotes

Mixed 50 µL of HA4 in EBY100 (0.98 x 106 cells) with 1 mL of FY250 (1.3 x 108)

Using an excess of FY250 to drive mating of all HA4; I want to be sure that I have diploids when I test yeast display ability.

Spun down 10 min at 3500 rpm, then dumped sup

Resuspended cells in 200 µL YPD, then dropped into center of a YPD plate

Put in 30 °C incubator face up at 12:00

4.6 Strains, Plasmids, and Oligonucleotides

Table 4-4. Strains used in this study

Name	Details	Source/Reference
EBY100	MAT**a** *ura3-52 trp1 leu2Δ1 his3Δ200 pep4:HIS3 prb1Δ1.6R can1 GAL* yeast display Aga system	
ysMDM100	EBY100 with *pep4:HIS3* replaced with *KanMX*	

Table 4-5. Plasmids used in this study

Name	Details	Source/Reference
pHRS-YD1B		
pHRS-YD2B		
pRS425	2µ *LEU2* pBIISK ori amp^R	ATCC #77106
pHRS-Cas2	pRS425 with HA4 BC loop	
pHRS-Sce1	*tetO$_2$*	
pHRS-Sce2	*tetO$_7$*	

Table 4-6. Oligonucleotides used in this study

Name	Sequence
MDM0231	GAGGTCGACGGTATCGATAAGCTTGATATCGTAGGGATAACAGGGTAATGTTCCGACCAAACTGGAAGTTGT
MDM231amp	GAGGTCGACGGTATCGATAAG
MDM0232	CCGACCAAACTGGAAGTTGTTGCTGCGACCCCGACTAGCCTGCTGATCAGCTGG
MDM0233	GTGAAACCGGTGGTAACTCCCCGGTTCAGGAATTCACTGT
MDM0234	GGAACAAAAGCTGGAGCTCCACCGCGGTGGCATTACCCTGTTATCCCTATGGTACAGTGAATTCCTGAACCGG
MDM234amp	GGAACAAAAGCTGGAGCTC
MDM240	CCGTATCACGTACGGTGAAAC
MDM0423	CCGTACGTGATACGGTAATAVNNVNNVNNVNNVNNVNNVNNVNNCCAGCTGATCAGCAGGCTAG
MDM0424	CCGTACGTGATACGGTAATAVNNVNNVNNVNNVNNVNNCCAGCTGATCAGCAGGCTAG
MDM0425	CCGTACGTGATACGGTAATAVNNVNNVNNVNNVNNVNNVNNCCAGCTGATCAGCAGGCTAG
MDM0426	CCGTACGTGATACGGTAATAVNNVNNVNNVNNVNNVNNVNNVNNVNNCCAGCTGATCAGCAGGCTAG
MDM0431	TGGTACAGTGAATTCCTGAACCGGGGAGTTACCACCGGTTTCACCGTACGTGATACGGTAATA
MDM0453	CCGTACGTGATACGGTAATARNNRNNRNNRNNRNNRNNRNNRNNCCAGCTGATCAGCAGGCTAG
MDM0454	CCGTACGTGATACGGTAATARNNRNNRNNRNNRNNRNNCCAGCTGATCAGCAGGCTAG
MDM0455	CCGTACGTGATACGGTAATARNNRNNRNNRNNRNNRNNRNNCCAGCTGATCAGCAGGCTAG
MDM0456	CCGTACGTGATACGGTAATARNNRNNRNNRNNRNNRNNRNNRNNRNNCCAGCTGATCAGCAGGCTAG

4.7 References

1. Weaver-Feldhaus, J. M. *et al.* Yeast mating for combinatorial Fab library generation and surface display. *FEBS Letters* **564**, 24–34 (2004).

2. Lou, J. *et al.* Affinity maturation of human botulinum neurotoxin antibodies by light chain shuffling via yeast mating. *Protein Engineering Design & Selection* **23**, 311–319 (2010).

3. Bosley, A. D. & Ostermeier, M. Mathematical expressions useful in the construction, description and evaluation of protein libraries. *Biomolecular Engineering* **22**, 57–61 (2005).

4. Boder, E. T. & Wittrup, K. D. Yeast surface display for screening combinatorial polypeptide libraries. *Nature biotechnology* **15**, 553–557 (1997).

5. Nobel, H., Pike, J., Lipke, P. & Kurjan, J. Genetics of a-agglutunin function in Saccharomyces cerevisiae. *Molecular and General Genetics MGG* **247**, 409–415 LA – English (1995).

6. Lipke, P. N. & Kurjan, J. Sexual agglutination in budding yeasts: structure, function, and regulation of adhesion glycoproteins. *Microbiological reviews* **56**, 180–194 (1992).

7. Wojcik, J. *et al.* A potent and highly specific FN3 monobody inhibitor of the Abl SH2 domain. *Nat Struct Mol Biol* **17**, 519–527 (2010).

8. Yang, J. *et al.* Dissecting Cooperative and Additive Binding Energetics in the Affinity Maturation Pathway of a Protein-Protein Interface . *Journal of Biological Chemistry* **278** , 50412–50421 (2003).

Chapter 5

Future Directions

5.1 Abstract

We are now at a very exciting point with our technology. After overcoming a number of technical obstacles, we are now in a position to do some really interesting experiments with the libraries.

5.2 Discussion

New libraries

Perhaps the most exciting new direction right now is just to build many different libraries, mate them, and see how they look on flow cytometry.

One particular route to building smart libraries is to utilize computational guidance[1–4]..

Characterization of Binding Hits

Conveniently, flow cytometry is an excellent way for measuring affinity of individual binders[5,6]. For binders in the nM affinity range, K_d values determined experimentally by SPR and flow cytometry are in excellent agreement[7]

SPR is a flexible and robust technique for the characterization of binding kinetics for an interaction between a protein and its ligand[8,9].

ITC is a technique that can determine K_d without the use of labels[10]. We have already begun using ITC with the SH2 model system (Figure 5-1).

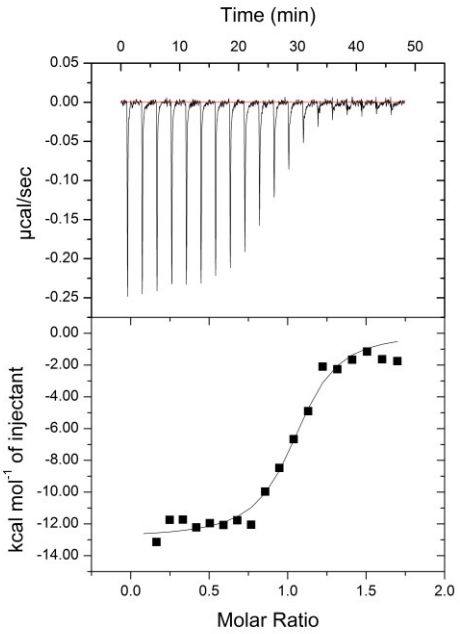


Figure 5-1. ITC data for the binding of HA4 ^{10}Fn3 to GST-SH2.

Re-run old libraries with new labeling

In Chapter 4, we discovered a method for FITC labeling that resulted in increased intensity signal by an order of magnitude. We realized that our old labeling method could only handle cell numbers less than 10^6, but we were running 10^8 cells in some of our libraries.

Specificity and additional SH2 domains

It would be interesting if we could use CDR walking technique to move from a high-affinity binder for one SH2 domain to another SH2 domain. Given the high degree of structural similarity between the SH2 domain family members, it is likely that a library developed for one will contain binders with some affinity for another. It would be really interesting to see if, while developing binders to multiple different SH2 domains in parallel, cross pollination from a library for target A into a library for target B could help the target B library move out of local minima in the search for more globally optimal solutions.

Improved Mutagenic Systems

Although our system was effective in bringing together diversity from two different libraries, searching sequence space is always a numbers game. All other things equal, better results will be obtained with larger library sizes. Going forward, a high-yield area for optimization of the technology will be in increasing the efficiency of the homologous recombination step.

One way to do this is to move the location of DNA double-strand break (DSB). In the current system, double-strand breaks are introduced on the flanks of the mutagenic cassette, which sits in the cassette plasmid. According to the current model, DNA resection creates single strands at the loose ends of the cut, which can then invade into the yeast display plasmid and initiate homologous recombination. However, the cassette plasmid is multi-copy, which means that there are roughly 20 copies per cell[11], whereas the yeast display plasmid is low-copy, with roughly 1-2 copies per cell[12]. Although the cassette in the cassette plasmid contains homology to the display plasmid, it clearly contains even greater homology, as it is in fact identical, to the 20 additional copies of the cassette plasmid within the same cell. Assuming no differences or biases between display and cassette plasmids in terms of intrinsic propensity for recombination, we would then expect that it is 20x more likely for a DSB within a given cassette plasmid to undergo the trivial, undesired homologous recombination event with an identical copy than to undergo the desired gene replacement repair with the display plasmid.

It is very reasonable to expect that the efficiency of homologous recombination would be improved by at least one order of magnitude (possibly more) if the site of DSB was moved to the middle of the ^{10}Fn3 gene in the yeast display plasmid, ideally in the framework region between the BC and FG loops. One way to do this might be through the use of introns. With an intron, you could put an SceI cut site directly between two loops at the DNA level, but it wouldn't be translated at the protein level.

Another alternative for mutagenesis is to generate the mutagenic cassettes by expressing mRNA at high copy and using a reverse transcriptase to generate DNA cassettes from the RNA, which could then participate in homologous recombination.

5.3 References

1. Pantazes, R. J. & Maranas, C. D. OptCDR: a general computational method for the design of antibody complementarity determining regions for targeted epitope binding . *Protein Engineering Design and Selection* **23** , 849–858 (2010).

2. Voigt, C. A., Mayo, S. L., Arnold, F. H. & Wang, Z.-G. Computationally focusing the directed evolution of proteins. *Journal of Cellular Biochemistry* **84**, 58–63 (2001).

3. Voigt, C. A., Mayo, S. L., Arnold, F. H. & Wang, Z. G. Computational method to reduce the search space for directed protein evolution. *Proceedings of the National Academy of Sciences, USA* **98**, 3778–3783 (2001).

4. Desjarlais, J. R. & Clarke, N. D. Computer search algorithms in protein modification and design. *Current Opinion in Structural Biology* **8**, 471–475 (1998).

5. Gera, N., Hussain, M. & Rao, B. M. Protein selection using yeast surface display. *Methods* (2012).at <http://www.sciencedirect.com/science/article/pii/S1046202312000709>

6. Boder, E. T. & Wittrup, K. D. Optimal Screening of Surface-Displayed Polypeptide Libraries. *Biotechnology Progress* **14**, 55–62 (1998).

7. Feldhaus, M. J. *et al.* Flow-cytometric isolation of human antibodies from a nonimmune Saccharomyces cerevisiae surface display library. *Nature biotechnology* **21**, 163–170 (2003).

8. Tanious, F. A., Nguyen, B. & Wilson, W. D. Biosensor-surface plasmon resonance methods for quantitative analysis of biomolecular interactions. *Methods in cell biology* **84**, 53–77 (2008).

9. McDonnell, J. M. Surface plasmon resonance: towards an understanding of the mechanisms of biological molecular recognition. *Current opinion in chemical biology* **5**, 572–7 (2001).

10. Velázquez Campoy, A. & Freire, E. ITC in the post-genomic era...? Priceless. *Biophysical Chemistry* **115**, 115–124 (2005).

11. Christianson, T. W., Sikorski, R. S., Dante, M., Shero, J. H. & Hieter, P. Multifunctional yeast high-copy-number shuttle vectors. *Gene* **110**, 119–122 (1992).

12. Futcher, B. & Carbon, J. Toxic effects of excess cloned centromeres. *Molecular and Cellular Biology* **6**, 2213–2222 (1986).

Appendix I

A word about serial processes affecting library complexity

It is important to inquire how my library complexity will be affected by lab bench procedures. A few papers have developed some basic mathematical descriptions for the manipulation of protein libraries[1–3]. An important rule of thumb, derived from Table 2 from Denault[1], is that when dealing with library complexities of 10^6, you need to work with a multiplicity of 20x ($2x10^7$ cells) in order for the probability of losing a unique clone to be 0.21% This is derived by using a Poisson distribution to represent the probability of loss of a unique clone from the library. This applies to any manipulation, even using a pipette to transfer cells from one tube to another. If instead, I only transfer 10^7 cells (10x of my complexity) the probability is ~100% that I have lost one of my unique clones. Thus, in all of the work described in this thesis relating to libraries of cells, I have made every effort to carry 20x my current estimate complexity through to the next step.

If I perform multiple library manipulations in a row, how is my overall library complexity affected? The following considerations are of a more "back-of-the-envelope" nature than the papers described above.

Due to the fact that our libraries are cells, and cells can be grown up and duplicated after each step, it can be readily shown that, to a first approximation, final complexity after the 3 manipulations is only a function of the single step with the lowest efficiency (i.e., the bottleneck). Let's say I start out with a library complexity of 10^8. And I perform three manipulations in a row (where a manipulation can be a mating, a selection, an induction step, etc.). Manipulation 1 has an efficiency of 10%, Manipulation 2 has an efficiency of 0.01%, and Manipulation 3 has an efficiency of 1%. If I took the cells output from one manipulation directly into the next, then I would simply multiply these all together. After the first, I'd have 10^7 cells, after the second, I'd have 10^3, and after the third, I'd only have 10 cells left.

Since I grow the cells to saturation after each step, I am really only limited by the number of cells I can physically grow. For most manipulations that are relevant to this work, the maximum number of yeast cells we can work with is 10^{10}. 100 mL can get sufficient oxygenation from being grown in a 500 mL flask (at the smallest). Let's say I fill my shaker with 6 1L flasks with 200 mL each. I can reasonably grow

$3.6x10^{10}$ cells, to the chagrin of my labmates who might also want to utilize the 30 °C shaker. This is about the ceiling of what is practical to work with on a non-industrial scale.

Now let me repeat the original sequence of manipulations. First, I amplify my 10^8 to 10^{10} by growing in culture. Now, on average, I have 100 copies of each clone from the 10^8 library. Then I undergo Manipulation 1, with efficiency of 10%. This take me to 10^9 cells; 10 out of the 100 copies is successfully output. I grow back up to 10^{10} cells, then undergo Manipulation 2, with efficiency of 0.01%. Now I've only got 10^6 cells successfully output, and I've lost 99% of my 10^8 sequence complexity. I grow up to 10^{10} again, so that I have 10^4 copies of each of my 10^6 clones now, and undergo Manipulation 3, with efficiency 1%, and my library complexity is unchanged since 1/100 doesn't make much of a dent in the 10^4 multiplicity for each unique clone.

Whenever the library size is at a high multiplicity of the complexity, these simple efficiency multiplications will hold. When I am dealing with low multiplicity, then statistical fluctuations will result in greater loss of complexity calculated by this simplistic method.

Appendix II

Programs for Mining Sequence Data

#0035 sequence processing.R

```
#This script harvests BC and FG loop sequences from primer 0035 reads of
#the FN3 protein in pYD (yeast display) plasmids.

#read in bioinformatics packages
library(seqinr)

#       set the directory that contains the sequencing data of interest
#dataDir <- "Research/Cornish Lab/Sequencing/2012-02-10 FN3 Library 96 well plate/10-
187894008_seq/a 0035"
#setwd(dataDir)

#       create an empty list that will be used to store raw DNA sequencing data
rawDNA <- list()
#       read all of the fasta format DNA sequences in the current directory into rawDNA
for (i in 1:length(dir())){
        tempSeq <- read.fasta(file = dir()[i], seqtype = "DNA")
        rawDNA[[i]] <- array(tempSeq[[1]])
}
rm(tempSeq)


#       the very beginning of the FN3 protein is marked by this nucleotide seq: GTTTCTTCT
#       create list that will hold all of the translated protein sequences
fn3Prot <- list()
```

```
#       for every raw data sequence that we have…
for (i in 1:length(rawDNA)) {
#       …pick out the numerical address at which that FN3 protein sequence begins
        FN3start <- regexpr("gtttcttct", c2s(rawDNA[[i]]))
        if (FN3start != -1){        #if FN3 start location was successfully found
                FN3end <- length(rawDNA[[i]])   #set an end-point
#       translate a protein sequence from the beginning of 10Fn3 to the end of the raw DNA read
                trans <- translate(rawDNA[[i]][FN3start:FN3end], frame = 0, sens = "F")
#       add this translation to the growing list of translations
                fn3Prot <- c(fn3Prot, list(trans))
        }
}
rm(FN3start, FN3end)

#Create lists to hold just the BC loop and FG loop sequences
bcSeq <- list()
fgSeq <- list()

for (i in 1:length(fn3Prot)) {
#       find the starting and ending positions of the bc loop in the sequence array
#       start: SLLISW
#       end: YYRITYG
        trans <- fn3Prot[[i]]
        bcStart <- regexpr("SLLISW", c2s(trans))
        bcEnd <- regexpr("YYRITY", c2s(trans))
#       if we have sequences for the two regions flanking the BC loop
        if(bcStart != -1 && bcEnd != -1)
#       Pull out the residue sequence and store it in bcSeq
```

```
                bcSeq <- c(bcSeq, list(trans[(bcStart+6):(bcEnd-1)]))


#       Do the same for the FG loop
        fgStart <- regexpr("TITVYA", c2s(trans))
        fgEnd <- regexpr("PISINY", c2s(trans))


        if(fgStart != -1 && fgEnd != -1)
                fgSeq <- c(fgSeq, list(trans[(fgStart+6):(fgEnd-1)]))
}
rm(fgStart, bcStart, fgEnd, bcEnd, trans, i)


#       reformat the structure of the data so that it can be saved as a CSV table
bcStr <- lapply(bcSeq, c2s)
bcStr <- unlist(bcStr)
fgStr <- lapply(fgSeq, c2s)
fgStr <- unlist(fgStr)


#       write data to CSV table
write.table(table(bcStr), file="bc table.csv", sep=",")
write.table(table(fgStr), file="fg table.csv", sep=",")
```

#0035 sequence processing.R

```
#This script harvests BC and FG loop sequences from primer 0032 reads of
#the FN3 protein in pYD (yeast display) plasmids.

#read in bioinformatics packages
library(seqinr)

#set the directory that contains the sequencing data of interest
#dataDir <- "Research/Cornish Lab/Sequencing/2012-02-10 FN3 Library 96 well plate/10-
187894008_seq/a 0032"
#setwd(dataDir)

#create a list with each element a raw DNA sequence read
rawDNA <- list()
#read DNA sequence in fasta format into a list called rawDNA
for (i in 1:length(dir())){
        tempSeq <- read.fasta(file = dir()[i], seqtype = "DNA")
        rawDNA[[i]] <- rev(comp(array(tempSeq[[1]]), ambiguous=TRUE))
}
rm(tempSeq)

#create list of translated protein sequences
#the very beginning of the FN3 protein: GTTTCTTCT
fn3Prot <- list()
for (i in 1:length(rawDNA)) {
```

```
	FN3start <- regexpr("gtttcttct", c2s(rawDNA[[i]]), fixed=TRUE)  #pick out location that FN3 DNA
sequence begins
	if (FN3start != -1){ #if FN3 start location was successfully found
		FN3end <- length(rawDNA[[i]])   #set an end-point
		trans <- translate(rawDNA[[i]][FN3start:FN3end], frame = 0, sens = "F") #translate a
protein sequence
		fn3Prot <- c(fn3Prot, list(trans))
	}
}
rm(FN3start, FN3end)

#Read out the BC loop sequences
bcSeq <- list()
fgSeq <- list()

for (i in 1:length(fn3Prot)) {
	#find the starting and ending positions of the
	#bc loop in the sequence array
	#start: SLLISW
	#end: YYRITYG
	trans <- fn3Prot[[i]]
	bcStart <- regexpr("SLLISW", c2s(trans))
	bcEnd <- regexpr("YYRITY", c2s(trans))

	if(bcStart != -1 && bcEnd != -1)
		bcSeq <- c(bcSeq, list(trans[(bcStart+6):(bcEnd-1)]))

	fgStart <- regexpr("TITVYA", c2s(trans))
```

```
        fgEnd <- regexpr("PISINY", c2s(trans))


        if(fgStart != -1 && fgEnd != -1)
                fgSeq <- c(fgSeq, list(trans[(fgStart+6):(fgEnd-1)]))
}
rm(fgStart, bcStart, fgEnd, bcEnd, trans, i)


bcStr <- lapply(bcSeq, c2s)
bcStr <- unlist(bcStr)
fgStr <- lapply(fgSeq, c2s)
fgStr <- unlist(fgStr)
write.table(table(bcStr), file="bc table.csv", sep=",")
write.table(table(fgStr), file="fg table.csv", sep=",")
```

References

1. Denault, M. & Pelletier, J. N. Protein Library Design and Screening. *Protein Engineering Protocols* **352**, 127–154 (2007).

2. Bosley, A. D. & Ostermeier, M. Mathematical expressions useful in the construction, description and evaluation of protein libraries. *Biomolecular Engineering* **22**, 57–61 (2005).

3. Kong, Y. Calculating complexity of large randomized libraries. *Journal of Theoretical Biology* **259**, 641–645 (2009).